

STIC-Biotech/ChemLib

61167

From: Pak, Michael
Sent: Wednesday, February 27, 2002 10:15 AM
To: STIC-Biotech/ChemLib
Subject: 08/816,011 sequence search

Sequence search - 2 month amendment
App. #: 08/816,011
Result format: Paper.
Title: potassium channels ...

Please search:

Search commercial and interference database.

SEQ ID NO:36 and 63.

Thanks,

Mike Pak

Michael Pak
Art Unit 1646
Mailbox: CM1, Rm. 10C00
Office: CM1, Rm. 10E13
703-305-7038

Michael Pak
USPTO
Art Unit 1646
CM1; Rm. 10E13
703-305-7038

Handwritten notes:
CIP of PCT/US 85/14364 10/25/95
CIP of 08/332,312 10/31/94 Pat # 5558028
only 99%
no support

RECEIVED
FEB 27 2002
STIC

if Contact:
Sheppard

Searcher: _____
Phone: tel: 308-4499
Location: _____
Date Picked Up: _____
Date Completed: 3/1/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 16:59:59 ; Search time 25.85 seconds
(without alignments)
2455.791 Million cell updates/sec

Title: US-08-816-011f-63

Perfect score: 2294

Sequence: 1 MVTHNSNTYAVVEHPRD.....LTGGLWPVVEHFDKPSIL 434

Scoring table: BLOSUM62

Gapop 10.0

Gapext 0.5

Searched: 473505 segs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing First 45 summaries

Database :

SPTREMBL17:*

1: sp-archaea:*

2: sp-bacteria:*

3: sp-fungi:*

4: sp-human:*

5: sp-invertebrate:*

6: sp-mammal:*

7: sp-mhc:*

8: sp-organelle:*

9: sp-phase:*

10: sp-plant:*

11: sp-rodent:*

12: sp-virus:*

13: sp-vertebrate:*

14: sp-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2265	98.7	434	5 Q19495	Q19495 caenorhabdi
2	802	35.0	418	5 Q9U318	Q9U318 caenorhabdi
3	716.5	31.2	461	5 Q23599	Q23599 caenorhabdi
4	706.5	30.8	450	5 Q93871	Q93871 caenorhabdi
5	645	28.1	458	5 Q9VMB6	Q9VMB6 drosophila
6	574.5	25.0	800	5 Q21145	Q21145 caenorhabdi
7	367.5	16.0	475	11 Q9UP65	Q9UP65 mus musculu
8	343.5	15.0	456	6 Q9UJY2	Q9UJY2 homo sapien
9	336.5	14.7	456	6 Q9BG85	Q9BG85 mus musculu
10	328.5	14.3	456	11 Q9JIT8	Q9JIT8 mus musculu
11	327.5	14.0	475	4 Q9BZD2	Q9BZD2 homo sapien
12	320.5	14.0	428	10 Q9M5X9	Q9M5X9 arabisidopsi
13	313.5	13.7	428	11 Q64603	Q64603 arabisidopsi
14	302.5	13.2	458	11 Q9JHF0	Q9JHF0 mus musculu
15	302.5	13.2	458	11 Q9DET8	Q9DET8 mus musculu
16	301.5	13.1	460	11 Q9JIM1	Q9JIM1 mus musculu
17	301.5	13.1	460	11 Q99K84	Q99K84 mus musculu
18	291.5	12.7	476	5 Q9NH21	Q9NH21 drosophila
19	261.5	11.4	397	4 Q9NUS9	Q9NUS9 homo sapien

Q9FWV1 arabidopsis
Q9VU20 drosophila
Q9VPP0 drosophila
Q9GTP5 crithidia f
Q93690 caenorhabdi
Q9BG84 oryctolagus
Q76343 leishmania
Q76269 leishmania
Q9PW12 homo sapien
Q9NR11 leishmania
Q20396 caenorhabdi
Q9NBV4 leishmania
Q9U763 trypanosoma
Q9SR64 arabidopsis
Q9M0Y3 arabidopsis
Q9TVQ1 toxoplasma
Q9M0Y2 arabidopsis
Q9Y010 trypanosoma
Q9NIH8 plasmodium
Q9CKV7 pasteurella
Q9Y0H9 trypanosoma
Q9GTP4 crithidia f
Q9B8V3 schistosoma
Q9G510 trigona ana
Q9NIH7 plasmodium
Q9SY87 arabidopsis

20 249.5 10.9 408 10 Q9FWV1
21 221.5 9.7 586 5 Q9VU20
22 217.5 9.5 404 5 Q9VPP0
23 197 8.6 497 5 Q9GTP5
24 196.5 8.6 513 5 Q93690
25 195.5 8.5 415 6 Q9BG84
26 195 8.5 491 5 Q76343
27 187 8.2 491 5 Q76269
28 185.5 8.1 285 4 Q9PW12
29 181 7.9 501 5 Q9NR11
30 180 7.8 143 5 Q20396
31 170 7.4 499 5 Q9NBV4
32 151 6.6 463 5 Q9U763
33 147 6.4 417 10 Q9SR64
34 144 6.3 418 10 Q9M0Y3
35 142 6.2 462 5 Q9TVQ1
36 138.5 6.0 418 10 Q9M0Y2
37 136 5.9 463 5 Q9Y010
38 135 5.9 422 5 Q9NIH8
39 135 5.9 487 2 Q9CKV7
40 133 5.8 463 5 Q9Y0H9
41 132.5 5.8 502 5 Q9GTP4
42 132 5.8 530 8 Q9B8V3
43 131 5.7 347 8 Q9G510
44 131 5.7 422 5 Q9NIH7
45 128.5 5.6 382 10 Q9SY87

ALIGNMENTS

RESULT 1

Q19495 PRELIMINARY; PRT; 434 AA.

AC Q19495;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SIMILARITY TO HNP36 PROTEIN.
GN F16H11.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX Wu X.;
RN [3]
RP Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U55376; AAA98003.1; -;
DR InterPro; IPR002259; DER_Nucleoside_tran.

DR Pfam; PF01733; Nucleoside_tran; 1.
DR ProDom; PD005103; DER_Nucleoside_tran; 1.
SQ SEQUENCE 434 AA; 49316 MW; 66947B8D51833965 CRC64;

Query Match 98.7%; Score 2265; DB 5; Length 434;
Best Local Similarity 99.1%; Pred. No. 3.4e-141;
Matches 430; Conservative 0; Mismatches 4; Indels 0; Caps 0;

QY 1 MVIINRSNTAVEQAEAPROKYNIVYLVILVGVLLPWNMFITIAPEYVYVWFKPDG 60
DB 1 MVIINRSNTAVEQAEAPROKYNIVYLVILVGVLLPWNMFITIAPEYVYVWFKPDG 60
QY 61 VETWYSKEFGSLTIGSQLPNASINVENFLIITAGPLIYRVEAPVCFNIVNLITILVI 120
DB 61 VETWYSKEFGSLTIGSQLPNASINVENFLIITAGPLIYRVEAPVCFNIVNLITILVI 120
QY 121 VLEPTEDSMGFWVTTLGMAISINFSNGLYENSIVYGVGGDFPHYIGALLIGNNICGLLI 180
DB 121 VLEPTEDSMGFWVTTLGMAISINFSNGLYENSIVYGVGGDFPHYIGALLIGNNICGLLI 180
QY 181 TVVKIGVYFLNDEPKLVAIVYFGISLVLLVCAIALEFFITKQDFYHYHOKGMEIREKA 240
DB 181 TVVKIGVYFLNDEPKLVAIVYFGISLVLLVCAIALEFFITKQDFYHYHOKGMEIREKA 240
QY 241 ETRDPSILMTTFTNCYGOLENNWFCFAVTLTIFPVMMVTVTRGDSGFLNKMSENDEI 300
DB 241 ETRDPSILMTTFTNCYGOLENNWFCFAVTLTIFPVMMVTVTRGDSGFLNKMSENDEI 300
QY 301 YTLTSLFVNLFNFAAGSIVASKIHWPTPRYLKFAILRALFPFFFCNRYVOTRAYPV 360
DB 301 YTLTSLFVNLFNFAAGSIVASKIHWPTPRYLKFAILRALFPFFFCNRYVOTRAYPV 360
QY 361 FFESTDIFVGGTAMSPSHGYSALAMGYTPNVVPSHSYRFAAQLSVCTLMVGLLTGGLW 420
DB 361 FFESTDIFVGGTAMSPSHGYSALAMGYTPNVVPSHSYRFAAQLSVCTLMVGLLTGGLW 420
QY 421 PVVIEHVDKPSIL 434
DB 421 AVVIEHVDKPSIL 434

RESULT 2
Q9U3L8 PRELIMINARY; PRT; 418 AA.
ID Q9U3L8;
AC Q9U3L8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE C47A4.2 PROTEIN.
GN C47A4.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Mortimore B.J.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Smalton N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans*;
RL Nature 368:32-38(1994).
DR EMBL; Z68303; CAA92642.1; -.
DR InterPro; IPR002259; DER_Nucleoside_tran.
DR Pfam; PF01733; Nucleoside_tran; 1.
DR ProDom; PD005103; DER_Nucleoside_tran; 1.
SQ SEQUENCE 418 AA; 46333 MW; 5CBDD27EE06BFC8D CRC64;

Query Match 35.0%; Score 802; DB 5; Length 418;
Best Local Similarity 37.6%; Pred. No. 2.8e-45;
Matches 156; Conservative 84; Mismatches 135; Indels 40; Caps 2;

QY 3 IINRSNTAYVEQAEAPROKYNIVYLVILVGVLLPWNMFITIAPEYVYVWFKPDGVE 62
DB 30 IVDGADSSDDEHNLIPEDKGRIVFWIILNGIGVLLPWNMFITIAPOYYDYVWFTVNGTA 89
QY 63 TWTYSKEFGSLTIGSQLPNASINVENFLIITAGPLIYRVEAPVCFNIVNLITILVI 122
DB 90 THVADSEQSGAMGVAVQPNLIVAIINVLIRPLLRVLAPLAFNLSLLIVIALVIFQ 149
QY 123 EPTEDSMGFWVTTLGMAISINFSNGLYENSIVYGVGGDFPHYIGALLIGNNICGLLI 182
DB 150 QPSDQARNWEYIVSLIIVMAMNASNGLYQNSFFGMAADFPKAYSNAVVICGTGTSV 209
QY 183 VKTGVYFLNDEPKLVAIVYFGISLVLLVCAIALEFFITKQDFYHYHOKGMEIREKAET 242
DB 210 LAIVATLAFSTQAEIVYFGISLVLLVCAIALEFFITKQDFYHYHOKGMEIREKAET 250
QY 243 DRSPSILMTTFTNCYGOLENNWFCFAVTLTIFPVMMVTVTRGDSGFLNKMSENDEIYT 302
DB 251 -----CWLCQICVFLVYFVSUSVPTVLVAGFQPGYTVPPN-----DVYA 289
QY 303 LLTSLFVNLFNFAAGSIVASKIHWPTPRYLKFAILRALFPFFFCNRYVOTRAYPVFF 362
DB 290 GIAVELNFNFFAAGVNAATFVTPGPRLLIVPCVIRLLFPFFEMESNYLPHSKRTMGVLF 349
QY 363 ESTDIFVGGIAMSFSHGYSALAMGYTPNVVPSHSYRFAAQLSVCTLMVGLLTG 417
DB 350 TNEWIFFFGNTLAFVTSYFSSGLMMYTPRVCPPEYSKLAGQVSALSILVILGITAG 404

RESULT 3
Q23599 PRELIMINARY; PRT; 461 AA.
ID Q23599;
AC Q23599;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ZK809.4 PROTEIN.
GN ZK809.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Dobson R.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-94150718; PubMed-7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Smalton N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans*;
RL Nature 368:32-38(1994).
DR EMBL; Z68303; CAA92642.1; -.
DR InterPro; IPR002259; DER_Nucleoside_tran.
DR Pfam; PF01733; Nucleoside_tran; 1.
DR ProDom; PD005103; DER_Nucleoside_tran; 1.
SQ SEQUENCE 461 AA; 52050 MW; CD0F115AF6A12A76 CRC64;

Query Match		31.2%; Score 716.5; DB 5; Length 461;
Best Local Similarity		35.8%; Pred. No. 1.2e-39;
Matches 163; Conservative 86; Mismatches 167; Indels 39; Gaps 9;		
QY	7 SNTYAYEQEA-FPRDKYNIYVWLVILVCGVLLPWNMFITIAPEYVYVWF-KPD	59
DB	11 NKTKKVEEESPEDEKGNLVFYIILHIGITLMPWNMLTISYDYFESYKMLANSTIDMD	70
QY	60 -----GVETWYKEFMGSLTIGSQLPNASINVFNFLIIAGPLIYRVFAPVCNIVNLTI	114
DB	71 TGVVTCYPIVYSSNFQSTIASQVNLNLLNFIIVKGGASRI--TVGLSIVAVCV	128
QY	115 ILILVILEPTEDSMGFWVTLMGMAISNFSNGLYENSIVYGGDPHPTIYIGALLIGNN	174
DB	129 ITTMMFIYVETSTWLTGFTLTITIIIVLNGANGVYQNSIFGLASELPFKYTNNAVIIGN	188
QY	175 ICGLLITVVKI----GVTYFLNDEPKLVAIVYFGISLVLLVCAIALFFITKQDFYHYHHQ	231
DB	189 LCQFTVTLTSMSTKAVTRNIDLR---SFAYFSIALITLVFCFISFHILKKORFYQYST	244
QY	232 KGMEIREK---AETDRSPSILWTTFTNCYQGFNVWFCFAVTLTIFF-VMTVTVTRGDS	287
DB	245 RAERQRKNKDEAVDSGKVANIATFKEAPPQQLINVLVFFVTLSPGVMVYKDEKKG	304
QY	288 G-----FLNKIMSENDEI---YTLTSLFVNLFAAIGSIVASKIHWPTTPRYLK	333
DB	305 GYVDFPLPQQFINSLEIHNREFSEKYFMDVTLFQNFVFAFGISVAGRKQWPAPNKLW	364
QY	334 FAILLRALIPFFFCNRYVOTRAYVPVFESTDIFVIGGIAISFSGHYLSALAMGTTPNV	393
DB	365 IPVYLRLLIYPIFFICNYLPETRSLSVPFFESTWLFVIIAASMSFGSGYTSGLAMYTSTK	424
QY	394 VPSHYSRFAAQLSVCTLVGLTGLGLWPVYVIEHFV 428	
DB	425 VDPKQAQVAGMAGFFLSIGVSLIFTWIKWV 459	
RESULT	4	
Q93871	PRELIMINARY; PRT; 450 AA.	
AC	Q93871; (TREMBlrel. 02, Created)	
DT	01-FEB-1997 (TREMBlrel. 02, Last sequence update)	
DT	01-FEB-1997 (TREMBlrel. 02, Last sequence update)	
DT	01-JUN-2001 (TREMBlrel. 17, Last annotation update)	
DE	K09A9.3 PROTEIN.	
GN	K09A9.3	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;	
OC	Rhabditidae; Peloderinae; Caenorhabditis.	
OX	NCBI_TaxID=6239;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Swinnburne J.;	
RL	Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=94150718; PubMed=7906398;	
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,	
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,	
RA	Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,	
RA	Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,	
RA	Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,	
RA	Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,	
RA	Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,	
RA	Snaldon N., Smith A., Sonnhammer E., Staden K., Sulston J.,	
RA	Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,	
RA	Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;	
RT	*2.2 Mb of contiguous nucleotide sequence from chromosome III of C.	
RT	elegans *;	
RL	Nature 368.32-38(1994).	
DR	EMBL; 279601; CAB01882.1; -	
DR	InterPro; IPRO02259; DER_Nucleoside_tran.	
DR	Pfam; PF01733; Nucleoside_tran; 1.	
DR	ProDom; PD005103; DER_Nucleoside_tran; 1.	
SQ	SEQUENCE 450 AA; 50495 MW; CF3973D5EB517909 CRG64;	
Query Match		30.8%; Score 706.5; DB 5; Length 450;
Best Local Similarity		36.0%; Pred. No. 5.5e-39;
Matches 157; Conservative 88; Mismatches 164; Indels 27; Gaps 8;		
QY	12 VEQEA-FPRDKYNIYVWLVILVCGVLLPWNMFITIAPEYVYVWF-----KPD	59
DB	21 VEETPEPEKGNLVFYIILHIGITLMPWNMLTISYDYFESYKMLANSTIDMDTGKVT	80
QY	60 GVTWYKEFMGSLTIGSQLPNASINVFNFLIIAGPLIYRVFAPVCNIVNLTIILV	119
DB	81 GDPVYSSNFQSTIASQVNLNLLNFIIVKGGASRI--TVGLSIVAVCVITMI	138
QY	120 IVLEPTEDSMGFWVTLMGMAISNFSNGLYENSIVYGGDPHPTIYIGALLIGNNCGLL	179
DB	139 FIYVETSTWLTGFTLTITIIIVLNGANGVYQNSIFGLASELPFKYTNNAVIIGNLCGT	198
QY	180 ITVVKI---GVTYFLNDEPKLVAIVYFGISLVLLVCAIALFFITKQDFYHYHHQGM	236
DB	199 VTLSSTKAVTRNIDLR---SFAYFSIALITLVFCFISFHILKKORFYQYSTRAERQ	254
QY	237 R---EKAETDRSPSILWTTFTNCYQGFNVWFCFAVTLTIFF-VMTVTVTRGDSGLNK	292
DB	255 RAKNEEAADNEGKMANVIATFKEAPPQQLINVLVFFVTLSPGVMVYKDEKKGTYDF	314
QY	293 INSENDEIYTLTSLFVNLFAAIGSIVASKIHWPTTPRYLKFAILLRALIPFFFCNRY	352
DB	315 PLPQN--YFMDVTLFQNFVFAFGISVAGRKQWPAPNKLWIPVYLRLLIYPIFFICNYL	372
QY	353 VQTRAYVPVFESTDIFVIGGIAISFSGHYLSALAMGTTPNVPSHYSRFAAQLSVCTLMV	412
DB	373 PETRLPVFESTWLFVIIAASMSFGSGYTSGLAMYTSTKTPDPSKAQVAGMAGFLIS	432
QY	413 GLTGLWPVYVIEHFV 428	
DB	433 GIVSGLIFTWIKWV 448	
RESULT	5	
Q9VMB6	PRELIMINARY; PRT; 458 AA.	
AC	Q9VMB6; (TREMBlrel. 13, Created)	
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)	
DT	01-JUN-2001 (TREMBlrel. 17, Last annotation update)	
DE	CG11045 PROTEIN.	
GN	BEST:LD04971 OR CG11045.	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=BERKELEY;	
RX	MEDLINE=20196006; PubMed=10731132;	
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Ananadis P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,	
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	
RA	Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,	
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,	
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	

DE GN EQUILBRATIVE NUCLEOSIDE TRANSPORTER 3.
OS ENT3.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL;
RA Hyde R.J., Cass C.E., Young J.D., Baldwin S.A.;
RT "The ENT family of eukaryote nucleoside and nucleobase transporters:
RT recent advances in the investigation of structure/function
RT relationships and the identification of novel isoforms.";
RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF326986; AAK00957.1;
RS
SQ SEQUENCE 475 AA; 51719 MW; F593D5D03C6CB52 CRC64;

Query Match 16.0%; Score 367.5; DB 11; Length 475;
Best Local Similarity 25.8%; Pred. No. 9.9e-17;
Matches 123; Conservative 83; Mismatches 194; Indels 77; Gaps 15;

Qy 7 SNTVAVQEAFA-----PRDKYNIVVMVLVILGVGFLLPNMFITTAPEYV 52
Db 21 SNHQEAQDEALLKLLDYPAQCLQPDGRFNKIIFFCLGIGGLPWNFVFT-AKEYWA 79
Qy 53 NYWFKPDGVETWYSKE-----FMGSLTIGSQLPNASINVFNLFIITAGPLIYRVF 102
Db 80 ---YKLRNCSSPASGEDPEDMDILNYFESYLAVASTVPSLLFLVANFLVNRVQVHVRVL 136
Qy 103 APVCFNIVNLTIILIVIVLEPTEDSMW----FWVTIGMATSNFNSGLYENSVYGVGG 159
Db 137 ASLSVSLAIFVMIVLVKV----DTSWTRGFFSLTIACMALIISSTIFNSSFVYGLTG 191
Qy 160 DFPTYICALLIGNICGLLTIVVKYIGVTFYFLNDEPKLVAIVYFISILVILLCAIALFF 219
Db 192 SFPMRNAALLSGGAMGG-TVSVALVLDLAASDVRDSTLAFFLMAVFLGLCMGYLL 250
Qy 220 ITKQDFVYHHQKWEIREKRAETDRPS---PS-----ILWTFETNCTY 258
Db 251 LSQLEYARYMYRPVAPRVFSGEDNPDSQAPSSAVAPASRVMTHTPLGPILKKT----- 305
Qy 259 GQLENVWFC-----PAVLTITPPVMVTVTRGDSGLFNKIMSENDEIYTLTSLFVNLEFA 314
Db 306 ---ASLGFCAVSLYFVTAFFIIPAISTNIQSMHKGTGSPWTS---KFFVPLTVFLLENFAD 359
Qy 315 AIGSIVASKIHWPTR--YLKFAILLRALFIPFFFCNRYQVTRAYPVFFESTDIF-VIG 371
Db 360 LCGRQVTAWIQVQPRSKLLPGLVFRCLFPLFLLCNYQPRSLTKVLFQS-DIYPLVF 418
Qy 372 GIAMSFSGYLSALAMGYTPNVPSHYSRFAAQLSVCTLMVGLTGLGLWPVVBIEHV 428
Db 419 TCLLGLSNGYLTSLVLYGPKIVPRELAETSVMVLFYMSVGLMGLSACAALLEHFI 475

RESULT 8
ID Q9UJY2 PRELIMINARY; PRT; 456 AA.
AC Q9UJY2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE EQUILBRATIVE NUCLEOSIDE TRANSPORTER 1 (SOLUTE CARRIER FAMILY 29
DE (NUCLEOSIDE TRANSPORTERS), MEMBER 1).
GN ENTI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Graham K.A., Coe I.R., Carpenter P., Baldwin S.A., Young J.D.,
RA Cass C.E.;

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF323951; AAK11605.1; -
SQ SEQUENCE 456 AA; 49858 MW; C6D55FC9D9D2A392 CRC64;

Query Match 14.7%; Score 336.5; DB 6; Length 456;
Best Local Similarity 24.9%; Pred. No. 1e-14;
Matches 115; Conservative 68; Mismatches 195; Indels 83; Gaps 13;
Qy 18 PRDXYNIYVWLIVLGVGVLPLPWNMTTIAPEYVYVWFKPDGVETWYSKEFMGS----- 72
Db 7 PRDSYHLVIGISFFILGLTLLPWNFFITAIP-YFOGRLAGANGAGALSTNHTGPTDAFN 65
Qy 73 ----LTTGSQLPNASINVENFLIAGPLIYRVFAPVCFNIVNLTILIVIVLEPTDS 128
Db 66 FNNVWTLSQLPPLLFTLLNSFLYQCIPAVRILG----SLLAMLLLFALTAALVKVDVS 121
Qy 129 MSWFFWVTLGMATSNFNSGLYNSVYGVGDFPHYTIAGLLIGNNICGL-----LITVV 183
Db 122 PGPPFSITMASAWFINSFCAVLQSLGQLGTMPTSTYNTLFLSGQLAGIPAALAMULTSM 181
Qy 184 KIGVTFYFLNDEPKLVAIVYF---GISLVLLVCAIALEFFITKQDFYHYHOK-----GM 234
Db 182 ASGV-----DAQTSALGYFLTPTCGVLSIVCYLSLPHL---EFARYLAKPSQAPTQ 232
Qy 235 EIREKAE-----TDRSPSILWTTFTNCTY 258
Db 233 ELETKAELLQCKEKNVPSSPKAALTSVDVDEKEPELEPAEPDGPKPSV-FVVFRIKI 291
Qy 259 GOLFNWVFCFAVTLTIFPVMVTVTRGDSGLFNKIMSENDEIYTLTSLFVFNFAIGS 318
Db 292 LTALCLVLVFTVLSVFPATAMVTSSTS-----PGKWSSEFFNPICCFLLFNVDMLGR 345
Qy 319 IVASKIHWP---TPRYLKFAITLRFALFPFFFCFCHYRVQTRAYPVFFESTDFIVGIGAMS 376
Db 346 SLTSTFLWVENSRLPLLVRLFLVPLFVLMCHVPOHAR-LPIIFPQAAFIIFMLLFA 404
Qy 377 FSHGYLSALAMGYTPNVVPSHYSFARAAOLSVCTLMVGLLTG 417
Db 405 VSNGLVSLTMCCLAPROVLAREVAGALMTFFFLAULGSLCG 445

RESULT 10
Q9JIT8
ID Q9JIT8 PRELIMINARY; PRT; 456 AA.
AC Q9JIT8
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE EQUILIBRATIVE NITROBENZYLTHIOMINOSINE-INSENSITIVE NUCLEOSIDE
DE TRANSPORTER ENT2.
GN SLC29A2 OR ENT2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD1; TISSUE=WHOLE BRAIN;
RA Kiss A., Farah K., Kim J., Garriock R., Drysdale T., Hammond J.R.;
RT "Molecular cloning and functional characterization of inhibitor-
sensitive and inhibitor-resistant equilibrative nucleoside transport
RT proteins from mouse brain."
RL Submitted (SEP-1999), to the EMBL/GenBank/DBJ databases.
DR EMBL; AF183397; AAF78477.1; -
DR MGD; MGI:11345278; SLC29a2.
DR InterPro; IPR002259; DER_Nucleoside_tran.
DR Pfam; PF01733; Nucleoside_tran; 1.
DR PRINTS; PR01130; DERENTRNSPT.
DR ProDom; PD005103; DER_Nucleoside_tran; 1.
SQ SEQUENCE 456 AA; 50255 MW; A9FE7CA037A4468F CRC64;

Query Match 14.3%; Score 328.5; DB 11; Length 456;
Best Local Similarity 24.2%; Pred. No. 3.4e-14;
Matches 113; Conservative 66; Mismatches 191; Indels 95; Gaps 14;
Qy 18 PRDXYNIYVWLIVLGVGVLPLPWNMTTIAPEYVYVWFKPDGVETWYSKEFMGS----- 72
Db 7 PRDSYHLVIGISFFILGLTLLPWNFFITAIP-----YFOGRLAGTNSAETMGNTHTSP 60
Qy 73 ----LTTGSQLPNASINVENFLIAGPLIYRVFAPVCFNIVNLTILIVIVLE 123
Db 61 TDTNFNNWVTLSQLPPLLFTLLNSFLYQCIPESVRLG----SLLAILLFLALTAALV 116
Qy 124 PTEDSMWFFWVTLGMATSNFNSGLYNSVYGVGDFPHYTIAGLLIGNNICGLITVV 183
Db 117 KVDLSPLGLFFSVTWASVWFINSFCAVLQSLGQLGTMPTSTYNTLFLSGQLAGIFAALA 176
Qy 184 KI-----GVTYFLNDEPKLVAIVYF---GISLVLLVCAIALEFFITKQDFYHYHOK----- 220
Db 177 MLMSLASGV-----DAQTSALGYFITPCVIGLLSIVCYLSLPHLKFARYLYTEKLSQAP 230
Qy 221 ----TKODFYHYHHOKGMEIR-----EKAETDRPSILWTTFTNCTY 259
Db 231 TQELTRKAEQLQADEKNGVPLSPQASPTLDLDPKEPEPEEPKPKPSVVFVFRKIWL 290
Qy 260 QLFNVWVFCFAVTLTIFPVM---MTVTTRGDSG---FLNKIMSENDEIYTLTSLFVFNLF 313
Db 291 TALCLVLVFTVLSVFPATAMVTSSTSNGKWLGFENPI-----CCFLFNVM 339
Qy 314 AAGISIVASKIHWP---TPRYLKFAITLRFALFPFFFCFCHYRVQTRAYPVFFESTDFIVI 370
Db 340 DWLGRSTSVFLWDEDSQQLPLLVCLRLFLVPLFVLMCHVPOHAR-LPIIFRQDAYFIT 398
Qy 371 GGIAFMSHGYLSALAMGYTPNVVPSHYSFARAAOLSVCTLMVGLLTG 417
Db 399 FMLLFAVNSGLVSLTMCCLAPROVLAREVAGALMTFFFLAULGSLCG 445

RESULT 11
Q9BZD2
ID Q9BZD2 PRELIMINARY; PRT; 475 AA.
AC Q9BZD2
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE EQUILIBRATIVE NUCLEOSIDE TRANSPORTER 3.
GN ENT3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA.
RA Hyde R.J., Cass C.E., Young J.D., Baldwin S.A.;
RT "The ENT family of eukaryote nucleoside and nucleobase transporters:
RT recent advances in the investigation of structure/function
RT relationships and the identification of novel isoforms."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF326987; AAK0958.1; -
SQ SEQUENCE 475 AA; 51874 MW; 59089BB8A5C75FA2C CRC64;

Query Match 14.3%; Score 327.5; DB 4; Length 475;
Best Local Similarity 24.4%; Pred. No. 4.1e-14;
Matches 116; Conservative 83; Mismatches 202; Indels 75; Gaps 15;
Qy 7 SNTYAVEQEAPEAF-----PRDKYNIYVWLIVLGVGVLPLPWNMTTIAPEYV 52
Db 21 SSSLRADQEALEKLLDRPPGLQRPDRFCGTIIFFSLGSLGLLPWNFFIT-AKEYWM 79
Qy 53 NYWFK-----PDGVETWYSKEFMGSITIGSQLPNASINVENFLIAGPLIYR 100
Db 80 ---FKLRNSSSPATGEDPEGSD---ILNYFESYLAVASTVPSMLCLVANFLNVRVAVHIR 134

Db	183	SGSVLSLLRLIKAVYPQDPDGLRKSANLYFAVGIVVMVICAIVFYNAIKLPIKFEHEE	242
QY	232	KGME--IREKAE-----TDRPSESILWTTFTNYGOLFNFWFCFVATLTIFPVMVTTR	284
Db	243	RKNEELIREKSEEGSLTGLAWRTLLWDIVTKVSHGCFGLVLLYMTLSIFFCYIT---	298
QY	285	GDGFLNKMISENDIYT--LLTSE-----LVFNLFAGTSGISVASKIHWPTPRYLKFAII	337
Db	299	-----EDVHSEULLTOMYPIILLIAAYNVFDLVGKCTVAFVMEDEKIAVGGSI	345
QY	338	LRALFIPEFFCFNRYQTRAYPVYFEFSTDIFVGGTAMSFSGHYLSALAMGYTPNVVPSH	397
Db	346	ARLLFPLFGCLHG-----PMLRTEIPVTTLTCLGLTNGYLYLSVLMILAPKSVPLR	399
QY	398	YSRFAQLSVCTLMVGLLTGGL--WPVVI	424
Db	400	HSSTAGIVTTFVLVVGGLASGVLTAWFWI	428

RESULT	13
ID	064603
PRT;	428 AA.
PRELIMINARY;	
AC	064603;
DT	01-AUG-1998
DT	(TREMBlrel. 07, Created)
DT	01-AUG-1998
DT	(TREMBlrel. 07, Last sequence update)
DT	01-JUN-2001
DT	(TREMBlrel. 17, Last annotation update)

Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidops.
OX NCBI_TaxID=3702;
RN [1]
RN
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G., Kwan A., Oji O.,
RA Liu S., Li J., Araujo R., Au M., Brendel V., Buehler E., Conway A.B.,
RA Conway A.B., Dewar K., Feng J., Kim C., Kurtz D., Li Y., Palm C.J.,
RA Shinn P., Sun H., Davis R.W., Eckert J.R., Federspiel N.A.,
RA Theologis A.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
RN [2]
RN
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
RN [3]
RN
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
RN [4]
RN
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AC003671; AAC19807.1; -;
DR InterPro; IPR002106; AA_TRNA_ligase_II.
DR InterPro; IPR002259; DER_Nucleoside_tran.
DR Prfam; PR01733; Nucleoside_tran; 1.
DR ProDom; PD005103; DER_Nucleoside_tran; 1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN1.
SQ SEQUENCE 428 AA; 46764 MW; B762E36C07516DB9 CRC64;

```

Query Match      13.7%  Score 313.5;  DB 10;  Length 428;
Best Local Similarity 23.2%  Pred. No. 3.1e-13;
Matches 100;  Conservative 84;  Mismatches 186;  Indels 61;  Gaps 11;

QY  18  PRKYNIVYWLVLIVLFGVLLPNMFMFTIAPEYKYNVWFKPDGVETYSKREFMGLSTIGS  77
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   35  PSUSYHEFAIYIFPLGVGFLPNNAFTIAYD--YFSYLYPSTAVDRIFAVIYM-----  85

```

```
QY 78 QLPNASINVENFLFLIA-----GLIYRVFAPVCFNIVNLTIIILVIVLEPTEDSMWF 132
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 86 ----LVALVCLFVIVVYFAHKSASFRINLGLLFFVIALLVVPVLDVYVKGQVGLYAG 140
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 133 FWTGLMATSFNSNGLYENSYVGVGGDFPHYTGALLIGNICGLLITVVKIGVTVFLN 192
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 141 FVDTSAVALSGGLDALMOGGLVAGEMPVYMOAVAGTAGSGVLVSLRLITRAVYP 200
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 193 DEP---KLVAIVFGISIVILLCAIAFFITTKODFYHHKQME--IREKAE-----T 242
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 201 QDPGLRKSANLPAVGIVVWVCAVYFNVVNAHKLVPVTKFHEERKEELIREKSEKGLT 260
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 243 DRPSPSLWTTFTTCYQGLFNVWFCEAVTLTIFPVMVTVTTGRDGSGLFNKIMSENDEIYT 302
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 GLAWRTTLWDIVTKVKGSHGFCVLLVYVWTLSPGYIT-----EDVHS 303
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 303 -LTSF-----LVFNLFPAAGISVASKIHWPTRYLKFALILRALFIPFFFCNRYVQT 355
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 304 ELLTDWYPIILIAAYNVFDLVGKCLTAVFMLEDEKIAVGSGIARLLFYPLFWGCLHG--- 360
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 356 RAYPVFFESTDIFVIGIAMSFSHYLSALAMGYTPNVVPSHYSRFAAQLSVCTLMVGLL 415
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 361 ---PMFLRTEIPVTLTCLGLGLTNGYLTSLVLMILAPKSVPLRHSETAGIVTVMFVVGLA 417
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 416 TGGI--WPVVI 424
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 418 SGSVIAFWVI 428
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
Q9JHF0
ID Q9JHF0 PRELIMINARY; PRT; 458 AA.
AC Q9JHF0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE EQUILIBRATIVE NITROBENZYLTHIOINOSINE-SENSITIVE NUCLEOSIDE TRANSPORTER
EN1.
DE "Molecular cloning of the full-length cDNAs encoding the
RT nitrobenzylthioinosine-sensitive and nitrobenzylthioinosine-resistant
RT equilibrative nucleoside transporters from mouse brain."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[2]
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=CD1; TISSUE=WHOLE BRAIN;
RA Hammond J.R., Kiss A.J.;
RT "Molecular cloning of the full-length cDNAs encoding the
RT nitrobenzylthioinosine-sensitive and nitrobenzylthioinosine-resistant
RT equilibrative nucleoside transporters from mouse brain."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[2]
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=129/SVJ1;
RA Choi D., Handa M., Young H.M., Messing R.O.;
RT "Genomic organization and expression of mouse equilibrative,
RT nitrobenzylthioinosine-sensitive nucleoside transporter 1 (es-ENT1)
RT gene."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=129/SVJ1;
RA Choi D.-S., Messing R.O.;
RT "Alternative splicing variants of equilibrative nucleoside transporter
RT 1."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF131212; AAF78452.1; -
DR EMBL; AF218255; AAF64036.2; -
DR MGI; MGI:1927073; SLC29a1.
DR InterPro; IPR002259; DER_Nucleoside_tran.
DR Pfam; PF01733; Nucleoside_tran; 1.
DR PRINTS; PRO1130; DERENTRNSPT.
```

```
DR ProDom; PD005103; DER_Nucleoside_tran; 1.
SQ SEQUENCE 458 AA; 50076 MW; EDE0FDEF5BF3E274 CRC64;

Query Match 13.2%; Score 302.5; DB 11; Length 458;
Best Local Similarity 22.5%; Pred. No. 1.7e-12;
Matches 104; Conservative 82; Mismatches 215; Indels 61; Gaps 13;

QY 18 PRKYNIVYLVILVGVGLLPNNMFITIAPEYVYVW----FKPDGVETWYSKEFMGS 72
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7 PQDYKAVLWIFFVGLGLTLLPNFMT-ATKFTNRLDSQNVSSDSCSTKALAD 65
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 LTIGSQLPNASINVENFLFLIAAGLIYRVFAPVCFN-----IVNLTIIIL 116
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 PTVALPARSSLSAIFNNVMTLCAMLPLLVF--TCLNSFLHQRIQSQRILGSLLAILLVF 123
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 117 ILVIVLEPTEDSMWFVWTLGMATSFNSNGLYENSYVGVGGDFPHYTGALLIGNIC 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 LVTAALVKVEMDALIFFVITMIKIVLINSFGAILQASLFLAGVLVPANYTAPIMSGOLA 183
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 GLLTIVVKIGVTVFLNDEPKLVAIVFGISIVILLCAIAFFITTKODFYHHKQME- 235
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 184 GFTSVAMI-CATASGSELSESAFGYFITAACAVVILAILCYLALPRTEFRHYLQLNAG 242
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 236 -----IRE-----KAETDRPSP-----SILMTTETNCYQGLFNVWFCEAVT 271
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 243 PAQETKLDLIREPKGRREESGVPGPSPTNRNQSIKAILKSICVPAL-SVCFIFTVT 301
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 272 LTIFPVMVTVTTGRDGSGLFNKIMSENDEIYTLTSLFVNLFAAGISVASKIHWPT-TP 329
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 302 IGLFP---AVTAEVSESIAG--TSPMKSYFIPVACFLNFVDFWGLRSLTAVCMWPQDS 356
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 330 RYLKFAILRALFIPFFFCNRYVQ---TRAYPVFFESTDIFVIGIAMSFSHYLSALA 386
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 RWLPVLVASRIVFIPLMLCNVRAHCGAQRHHFVFKHDAWFAFMAAFSAFNGYLASLC 416
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 387 MGYTPNVVPSHYSRFAAQLSVCTLMVGLLGLVGPVVIHFV 428
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 417 MCFGPKKVKPAEATAGNIMSFELCLGLALGAVLSFLRLALV 458
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
Q9DBT8
ID Q9DBT8 PRELIMINARY; PRT; 458 AA.
AC Q9DBT8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 1200014D21RIK PROTEIN.
GN 1200014D21RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=LUNG;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto K.-F.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
```

Search completed: February 27, 2002, 17:03:21
Job time: 202 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2002, 23:22:38 ; Search time 1347.44 Seconds
(without alignments)
16993.770 Million cell updates/sec

Title: US-08-816-011f-36
Perfect score: 1388
Sequence: 1 atgtgtaataatgaacgatc.....ttattaaaaaaaaaaaaa 1388

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pal:*
- 7: gb_ph:*
- 8: gb_pi:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_om:*
- 20: em_or:*
- 21: em_ov:*
- 22: em_pat:*
- 23: em_ph:*
- 24: em_pi:*
- 25: em_ro:*
- 26: em_sts:*
- 27: em_sy:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htgo_hum:*
- 31: em_htgo_inv:*
- 32: em_htgo_rod:*
- 33: em_htg_hum:*
- 34: em_htg_inv:*
- 35: em_htg_rod:*
- 36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	101.2	72.9	31049	3	U55376	U55376 Caenorhabdi
c 2	90.8	6.5	13429	2	AC019610	AC019610 Drosophi
c 3	90.8	6.5	64674	3	AC004247	AC004247 Drosophi
c 4	90.8	6.5	174376	3	AC092230	AC092230 Drosophi
c 5	90.8	6.5	259464	3	AE003614	AE003614 Drosophi
c 6	84.8	6.1	31853	3	CEFA4B12	282263 Caenorhabdi
c 7	71.2	5.1	40664	3	CEFA4B12	282263 Caenorhabdi
c 8	70.8	5.1	25560	3	CEZK809	268303 Caenorhabdi
c 9	68	4.9	7218	6	166494	166494 Sequence 14
c 10	65.8	4.7	34017	3	CEK09A9	279601 Caenorhabdi
c 11	47.8	3.4	253305	3	PFMAL3P7	AL034559 Plasmodiu
c 12	44.6	3.2	3885	3	AF151733	AF151733 Dictyoste
c 13	44.6	3.2	3985	3	AF151111	AF151111 Dictyoste
c 14	44.6	3.2	110211	8	AC005170	AC005170 Arabidops
c 15	43.8	3.2	366	8	SCOR17	X59536 S.cerevisia
c 16	43.8	3.2	2468	8	YSCMTORIK	M11449 Yeast mitoc
c 17	43.8	3.2	21153	8	YSCMTG13	L13697 Saccharomyc
c 18	43.8	3.2	85779	8	SCE011856	AJ011856 Saccharom
c 19	43.8	3.2	112576	2	AC011250	AC011250 Drosophi
c 20	43.8	3.2	184621	3	AC007440	AC007440 Drosophi
c 21	43.8	3.2	262731	3	AE003823	AE003823 Drosophi
c 22	43.4	3.1	77835	2	PFMAL13P2_3	Continuation (4 of
c 23	43	3.1	1141	6	AX083744	AX083744 Sequence
c 24	43	3.1	33637	3	CEY10G11	AL161711 Caenorhab
c 25	42.8	3.1	175737	9	HS1033B10	AL031228 Human DNA
c 26	42.4	3.1	2175	3	AF139082	AF139082 Haematobi
c 27	42.4	3.1	321003	2	PFMAL4P3	AL035476 Plasmodiu
c 28	42.2	3.0	104992	2	AC005504	AC005504 Plasmodiu
c 29	42.2	3.0	169546	2	AC004157	AC004157 Plasmodiu
c 30	42	3.0	23269	8	SPAC1856	AL391783 S.pombe c
c 31	42	3.0	23269	8	SPAC1856	AL391783 S.pombe c
c 32	41.6	3.0	175438	2	AC084352	AC084352 Homo sapi
c 33	41.6	3.0	321003	2	PFMAL4P3	AL035476 Plasmodiu
c 34	41.2	3.0	64248	2	AC017951	AC017951 Drosophi
c 35	41.2	3.0	84055	3	AC004439	AC004439 Drosophi
c 36	41.2	3.0	170868	3	AC092221	AC092221 Drosophi
c 37	41.2	3.0	306352	3	AE003586	AE003586 Drosophi
c 38	41	3.0	315	3	ECA270225	AJ270225 Entodiniu
c 39	40.8	2.9	178586	2	AC073485	AC073485 Homo sapi
c 40	40.8	2.9	191122	9	AC012494	AC012494 Homo sapi
c 41	40.6	2.9	14923	3	AE001394	AE001394 Plasmodiu
c 42	40.6	2.9	88512	8	F21H2	AC007894 Arabidops
c 43	40.6	2.9	110619	8	AC018460	AC018460 Arabidops
c 44	40.6	2.9	199551	2	AC006281	AC006281 Plasmodiu
c 45	40.4	2.9	2206	8	SCYK072C	Z28297 S.cerevisia

ALIGNMENTS

RESULT 1	U55376/c	31049 bp	DNA	INV	09-AUG-2001
LOCUS	U55376	Caenorhabditis elegans cosmid F16H11, complete sequence.			
DEFINITION	U55376				
ACCESSION	U55376.1	GI:1280130			
VERSION	HTG.				
KEYWORDS	Caenorhabditis elegans.				
SOURCE	Caenorhabditis elegans				
ORGANISM	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;				
REFERENCE	Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.				
AUTHORS	1 (bases 1 to 31049)				
TITLE	The C. elegans Sequencing Consortium.				
JOURNAL	Genome sequence of the nematode C. elegans: a platform for				
MEDLINE	investigating biology. The C. elegans Sequencing Consortium				
REFERENCE	Science. 282 (5396), 2012-2018 (1998)				
AUTHORS	99069613				
TITLE	2 (bases 1 to 31049)				
	Wu.X.				
	The sequence of C. elegans cosmid F16H11				

JOURNAL
REFERENCE 3 (bases 1 to 31049)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL
REFERENCE 4 (bases 1 to 31049)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (18-APR-1996) Robert Waterston
AUTHORS (bases 1 to 31049)
JOURNAL Waterston,R.
TITLE Direct Submission
COMMENT Submitted (09-AUG-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
Submitted by: Genome Sequencing Center
Department of Genetics, Washington University
St. Louis , MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RO, England
email: rw@ematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

NEIGHBORING COSMID INFORMATION

The 5' cosmid is C46C11, 200 bp overlap; 3' cosmid is K02G10, 2500 bp overlap. Actual start of this cosmid is at base position 1 of CELF16H11; actual end is at 10112 of CELK02G10.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES
source 1. .31049
Location/Qualifiers
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="X"
/clone="F16H11"
complement(1239. .2839)
/gene="F16H11.3"
complement(join(1239. .1368,1421. .1617,1669. .1878, 1932. .2131,2180. .2466,2516. .2649,2693. .2839))
/note="weak similarity to HNP36 protein"
/codon_start=1
/product="Hypothetical protein F16H11.3"
/protein_id="AAA98003.1"
/db_xref="GI:1280131"
/translation="MVIIINRNTAYVEQEAFFPRDKYNIVVILVGFVGLLPNNF1
TIAPEYVNVKPDGVEWYKEFMGSLT IASQLPNASINVENFLIITAGPLIYRVF
APVCENIVLTILIVILEPTEDSMFFWVTLGMATISINSGLYSVYGVAD
PPTYTGALLIGNICGLITVYKIGVTFYFNDPEKLVAIYVIGISILVILVCAIALF
FITQDFYHHQKMEIREKAETDRPSPSLMTITNCTGQLFNWVFCFVAVLTIPP
VMVTTRGDSGLNKMEINDEIYTLTSLFVNLFAIGSIVASIKHWPTPYRIKE
AIIILRALFTPFPPFCNRYQTRAYPVFFESTDIFVIGGIAMSESHGYLSALAMGYTPN
VVPVSHYRFAAQLSVCTLMVGLLTGGLWAVVIEHFVDKPSIL"
complement(7137. .7683)
/gene="F16H11.4"
/note="ceh-1"

CDS
complement(join(7137. .7253,7307. .7456,7505. .7573, 7621. .7683))
/gene="F16H11.4"
/note="C. elegans unknown homeobox gene ceh-1 (GB:X52810)"
/codon_start=1
/product="Hypothetical protein F16H11.4"
/protein_id="AAA98004.1"
/db_xref="GI:1280132"
/translation="MRRARTAYTEQLVALENFKTSRYLSVVERLNLAQLQQLSETQ
VKTFONRRTKKKHNPQDANTPOTPPSSDETQIQIPILPANPITSSSSLLLPITSP
ANSVQLQTSIPLTLENLNLQILMPQNVGFN"
complement(10470. .14006)
/gene="F16H11.5"
/note="nhr-45"
complement(join(10470. .10481,10529. .10624,10701. .10861, 10907. .11025,11086. .11396,11509. .11613,12150. .12431, 13147. .13430,13566. .13631,13796. .14006))
/gene="F16H11.5"
/note="contains a zinc finger of the C4-type; coded for by the following C. elegans cDNAs: yk347g3.3, yk347g3.5"
/codon_start=1
/product="Hypothetical protein F16H11.5"
/protein_id="AAA98005.1"
/db_xref="GI:1280133"
/translation="MTTFHSINLTFFSETTSKISMSCLVCETAHQHGFIRCCRAC
AAPFRRLTMNLRYKCFDRKCEVSNFKRYSCRCRYEKCVRYGMRKDNLTSPVSVVQ
NKSETEKDSGNETSDIGSHPSHSLSYTCRPPSFQQQONQNDPPLLMNGNCTVQ
EVASDAYQNSDMHOLFSEFHRTYIEEVGRAVMPPTPFOQPIIHPQPFQYQVTDLLS
TDRSSATPSSETQYNOVQEYNNESGNQNMFAAQOATODLHNVNQNVLTPID
AMFGPPDAFCQLPDILQLCQALLAYREHNKQWQDQKNIENVPDLMENFNRNHYI
EHIHARFMSIRVFAQLPQDKWIIIFKHFWTRFYELDRFCATCQRLGYNLTERGTL
LNIHIFGIVSVKLEISMDQATNRFNLKSMDFRFLIFINPFKKLQPTTELYM
MMSIMSVSNLPGITDTRDISKVELRLAEDLHTYAEQYDNNPNYAGRITRLSSI
ASAVDETFRKRDQSQVSKTFNFKNDFEFLDTDFPCLN"
16440. .16613
/gene="F16H11.2"
16440. .16613
/gene="F16H11.2"
/codon_start=1
/product="Hypothetical protein F16H11.2"
/protein_id="AAA98006.1"
/db_xref="GI:1280134"
/translation="MTQDRDTEERERERERERERERERERERERERERERE
REREREKESNRKR"
21734. .26707
/gene="F16H11.1"
join(21734. .21787,21843. .21909,21958. .22079,24004. .24109, 24156. .24295,24739. .24809,24855. .24966,25036. .25127, 25173. .25257,25305. .25397,25441. .25536,25584. .25751, 26297. .26394,26451. .26593,26649. .26707)
/gene="F16H11.1"
/note="similar to melibiose carrier protein (thiomethylgalactoside permease II); coded for by the following C. elegans cDNAs: cm01e2, cm21e6"
/codon_start=1
/product="Hypothetical protein F16H11.1"
/protein_id="AAA98007.1"
/db_xref="GI:1280135"
/translation="MTDETEGDVVPVRSDDIVIPRRSIRPDHDSFTKKIAFGIGH
FYNDLCASMWFTYFIYMEVLSKFKQSRAGMLMIGVTDASTPLVGLFSDSNILPA
CFDKIGRRMSWHLIGTVLSLSPFMIFNCKFCICKSTISWKLKLVFVFMVTFQFNA
SVOISHLALIPELSSVPASRATMNSURAFYVIANLSVYPALAWLSESTGHMISGW
DFSHFACGLVVLGVITAVFYAFYREPTNYRRESLNSFSSDASELVNHTWISGW
GHQVFOYALLNLSRLYINISQVYFPFTMTQPKYKVALILPVAVALSSFSVSNV
NSLPVYSKSKLTLYSFLGASGLSCAVNMLDLPQWKYVALAVGICGAOAILTSL
ITADLINKTESCAFYVYGMSPFDKLSNGTAYOLIELWTPAYDALKPEHVAIFYRRV
MWPVGTCLVLAFLVLSLAPFKIGERRARPEQAINEDQDDIYPIVEI"
BASE COUNT 10124 c 5591 g 10234 t
ORIGIN

Query Match 72.9%; Score 1012; DB 3; Length 31049;
Best Local Similarity 82.0%; Pred. No. 3.3e-251;

Matches 1371; Conservative 0; Mismatches 5; Indels 296; Gaps 6;

Qy	1	atgtaataatcaaccgagtcgaacacccatcgcttgaagcaggaagcattccaaagaac	60
Db	2839	ATGGTAATAATCAACCGATCGAACACCTTATGCCGTTGAGCAGGAGCAATTTCCRAGAGAC	2780
Qy	61	aagtacaatttctactgctcgtcgtcattcttctgttgattcggagttctctcgcattg	120
Db	2779	AAGTACAATAATGCTACTGCTGCTCATTTCTGTTGGATTGCGAGTCTCTTCCCATGG	2720
Qy	121	aatatgttcattactatgcgcctcg-----	145
Db	2719	AATATGTTTCATTACTATCGCCCTGAGGTTGCAACTAGATAATTTTTCACATAATAGATT	2660
Qy	146	-----agtatgtgaattatctgttcaaacggatgctggtgagacatgattc	197
Db	2659	CCTGTTTCAGTATATGTAATATGTTTCAACCGGATGGGTTGGAGACATGGTATTC	2600
Qy	198	gaaagaattcatgggatctttgagattggctcacaacttccaaacgcaagcattaatgt	257
Db	2599	GAAGAATTCAATGGGATCTTTGAGGATTGCTCACAACCTTCCAAACGCAAGCATTAAATGT	2540
Qy	258	tttcaacctgttcctcattatgc-----	281
Db	2539	TTTCAACCTGTTCCCTCATTTATTTGCGTAAGTTACAGTTCTTTAAACTCTCAAAATCAAT	2480
Qy	282	-----tqgtccctgatctaccgctcttctgcccgttttcttcaacatcg	328
Db	2479	CAATATTTTTTAGTGGTCCCTGATCTACCGGCTTTGCTCCGGTTTGTCTCAACATCG	2420
Qy	329	tcaacctgacaatcatctcatcctcgtcattgttctggagcccactgaagattccattgt	388
Db	2419	TCAACCTGACAATCATCTCATCTCTCGTCATTTCTGAGGCCACCTGAAGATTCCATGT	2360
Qy	389	cctggttttctggtaactcttggaaatggcgaacttcaatcaatttttagaatgggctat	448
Db	2359	CCTGTTTTCTGGTAACCTCTGGAATGGCGACTTCAATCAATTTTAGCAATGGGCTAT	2300
Qy	449	atgaaactcggtttatgagttggtgagatttccgcacacactacattggcgctctct	508
Db	2299	ATGAAACTCGGTTATGAGATTTTGGCGATTTTCCGCACACCTACATTGGCGCTCTCT	2240
Qy	509	tgattggaaacaacatttcgaggtatgctgataacggttgtgaaatcggagtgacctatt	568
Db	2239	TGATTGGAACAACATTTGCGGATGCTGATACGGTTGTGAAATCGGAGTGACCTATT	2180
Qy	569	-----tctggaatgatg	580
Db	2179	GTAAGTAATTTTCAATTTATTTGCTTTTAAACATATAACATTTTATGTTCTGAATGATG	2120
Qy	581	agcctaaacttctgcaatcgtctatttccggcatatcgttggatccttctggtggtg	640
Db	2119	AGCCTAAACTTGTGCAATGCTATTTTCGGCATATCGTTGGTATCCTTCTGTTGCTGTG	2060
Qy	641	caattgcaactttctttatcaaaagcaagattttaccacatatcaccatcaaaaagaa	700
Db	2059	CAATTGCACTTTCTTTATCAAAAGCAAGATTTCTACCACTATCACCATCAAAAAGGAA	2000
Qy	701	tggaaattcggaaagcggaacccgacagacgctctccatcatttttggaccacat	760
Db	1999	TGGAAATTTCGGGAAAGGCGGAAACCGAGACCGCTCCCATTCCTTTTGGACCCAT	1940
Qy	761	tcacaac-----	768
Db	1939	TCACAAACGCTGAGTTTTTAAATCAATTTAGCTTCAAAATAATATGTGTGACTAATTTCA	1880
Qy	769	tgttatgggcaactcttcaatgtttgtctgttccgttactcttcaaatcttccc	827
Db	1879	GTGTTATGGCAACTCTCAATGTTTGTGTTCTGCTTCCGCTTACTCTCAACAATCTTCCC	1820
Qy	828	tgttatgatgaccttaccactcgtggagattccggttctcctaacaacaaatatatgtcga	887
Db	1819	TGTTATGATGCCGTTACCACTCGTGGAGATTCCGGCTTCCTAAACAAATATATGCTCTGA	1760

RESULT	2
AC019610	
LOCUS	AC019610 13429 bp DNA HTG 03-JAN-2000
DEFINITION	Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in Ordered pieces.
ACCESSION	AC019610
VERSION	AC019610.1 Gi:6665287
KEYWORDS	HTG; HTGS_PHASE2.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS	Adams M. and Venter, J.C.
TITLE	Direct Submission
JOURNAL	Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
COMMENT	This sequence was identified as CDM:10210382 by the submitter. For more information on this record e-mail to fly@celera.com. * NOTE: This is a 'working draft' sequence. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved.
FEATURES	Location/Qualifiers
source	1. 13429
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
BASE COUNT	3597 a 3053 c 2937 g 3842 t
ORIGIN	

Query Match 6.5%; Score 90.8; DB 2; Length 13429;
 Best Local Similarity 45.4%; Pred. No. 1.2e-12;
 Matches 556; Conservative 0; Mismatches 632; Indels 36; Gaps 5;

y	47	catttccaagaacaagtaacaatattgtctactggtcgtcgtcattcttggattcggag	106
b			
b	7103	CAGGCCAAAGATAAATTCCTTATCGTGCTCTTTATATCTCGTGTGCACGGCGTGGCA	7162
y	107	tctcttgccttgaataatgtctactactatcgcgc-----tgagattatgtgaatt	160
b			
b	7163	CGTGATGCCATGGAACATGTTTCATCAGCGCAAGTCCTATTTCGAGGACTTCAAGTTCG	7222
y	161	attggttcaaacggggtggtggtgagacatggtattcgaagaatcattggatcttga	220
b			
b	7223	GACCAAAACACACCGTGGCCACGGAGTAGTGACCGCACCCCATTTTCGAAAAATATGG	7282
y	221	cgattggtctcaaaccttccaaacgcaagcatcaatgttttcaacctgtctccattatg	280
b			
b	7283	GCATTGCCCTCGCAGATTCCTAACCTGGTCTTCAACTGGGTGAACATCTTGTCAACTTGG	7342
y	281	ctggtcccttgatctacgcgcgttttggctccgggttggcttcaacatcgtcaaacatgacaa	340
b			
b	7343	CGCGCAGACCTGACACCCGAAATCGTCTAGACGATCAICTTTCGAAATGGTTCATCTCTGCTGG	7402
y	341	tcaattcctctcgtctcatgttcttgagccacgtgaagatcccatgctcctgtgtttct	400
b			
b	7403	TTACCATTAATTTGGCCATGCTCG-----ACTCTCCGACGTGGCGGGCTTCTTCTCT	7456
y	401	gggtcaactcttggaatggtgcgaattcaatcaatttagcaatgggtatataaacaactcgg	460
b			
b	7457	GGACCACCACTGTGTGCATGTGTGTTTAAACGTGTGCAAGGCCATCTACCAAGAACACCA	7516
y	461	tttatggattggtggcgattttccgcacacatcatgtggcgctctcttgattggaaca	520
b			
b	7517	TATACCGAATAGTGGCATCGTTCGCCATCAAGTACACAGGCCCGCTGCTGGGCTCCA	7576
y	521	acatttgcggtattgtgataaaacgttggaaaatcggagtgacctattttctgaaatg	580
b			
b	7577	ACATCAGCGGCTGCTTCACACCGCCGATGGCTTGATATGCGGTGAGATCTTCTCGTCCA	7636
y	581	agcctaaactgttgcgaactcgtctattcttggcatatcgttggatcctcttggtgtg	640
b			
b	7637	AG---AGGACTTGGGGCATCTACTACTCTGTCAGCCGCATCCTACTGTGTGCTCTGCT	7693
y	641	caattgcacttttcttatacaaaagcaagatttctaccactatcaccatcaaaaaggaa	700
b			
b	7694	TCGACACGTACTTTGGGCTGCCGCTGAACAAGTCTTCCGGCATTACGAGACCATTAGCC	7753
y	701	tggaaattcgcgaaaggcggaaacgcagacgcgtctctcaatccttcttggaccacat	760
b			
b	7754	GGAGCAGCGAAAAGAAATCTGATTTCTAAGGCCACAGTCAAGCTGCGCTTACTGGCAGATCT	7813
y	761	tccaaactgttatgggcaactcttcaatgttttcttctgtcttgcgttactctcaaca	820
b			
b	7814	TCAAGAAACAGCGCCCCAACATATTCAACATCTTCTCAGCTTTTTCGTACGTGTGCG	7873
y	821	tcttccctgttatgatagcggttaccactcgttggagatcccggtctctcaacaaatta	880
b			
b	7874	TTTTTCCGGCGATTTCAGTTCGAAGTGCACCGTCCGACCCCAAAATTCGTA-----	7923
y	881	tgtctgaacagatgaaatctacacttctgcacaaagtttctcgtctcctaattgttcg	940
b			
b	7924	-----GTAGCCCGGACTACTTTTACATGGTCACTTGCTTCGCGACCTTTTACGTCTTCG	7978
y	941	ctgcgattggatcccatagttgcttccaagattcactggcgcgacaccccggttacctcaaat	1000
b			
b	7979	CAATGCTGGGCAGTTTAAACCATCATCGTGGGTCAATGCGCGGTCGAGGTTCTCTGGG	8038
y	1001	tgtgcaataacttgctgctcttcttcaatcctaattcttcttcttctgaacta-----tc	1054
b			
b	8039	TGCCAGTGTGCTGGCGTTCGGCGTTTCATCCCTCTGTTGTGATGTGCAACTACGTTCGCG	8098

Qy	1055	gtgtccagacgctgcttatacctgttttctttagtcctactgacatttttttgatgggtg	1114
Db	8099	CGGACTGGGTGGCTCATTTGGCCGTATTTCATCGAGAACGATTGGTGTACTCGGGGCATCG	8158
Qy	1115	gaattgcacatctcttttaccatgatataacctcaggcgctctggcaatggatatcacacitccaa	1174
Db	8159	GCATCGCAATGGCCTACAGCTCCGGGTATCTCAGCTCTCTGGGCATGATCTACGCCCCGC	8218
Qy	1175	acgttgtgccattctcacactactacaagtattgcgcctcagctttccgttttcacactcttatgg	1234
Db	8219	AGACGGTCGACACCAAATACCAGACCACTGCCCGGAATGTACGCAAGCTGCCATGCTGATCA	8278
Qy	1235	tggccttctcacggtgctgtg	1258
Db	8279	CGGGCATCTTCTCCGGAGTGCTGT	8302

RESULT	3
AC004247/c	
LOCUS	AC004247 64674 bp DNA INV 01-MAR-1998
DEFINITION	Drosophila melanogaster (P1 DS07773 (D154)) DNA sequence, complete sequence.
ACCESSION	AC004247 AC003251 AC003249 AC003250 AC003252 AC003248 AC003247
VERSION	AC003761
KEYWORDS	HTG.
SOURCE	Drosophila melanogaster (Subclones in Sac from P1 clone DS07773 (D154)) DNA.

ORGANISM Drosophila melanogaster (Linn.)
 Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta;
 Pterygota: Neoptera: Endopterygota; Diptera: Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 64674).
 REFERENCE
 AUTHORS Celniker, S.E., Aghavani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
 Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A.,
 Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B.,
 Lomoton, M.A., Mazda, P., Mok, M.S., Moshrefi, A.R., Moshrefi, M.,
 Nixon, K., Pacle, J.M., Park, S., Pfeiffer, B., Punch, D., Snir, E.,
 Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R., Zieran, L.L.
 and Kimmel, B.

TITLE Sequencing of *Drosophila* chromosome 2L, region 26F1-26F4
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 46474)
AUTHORS Celniker, S.E., Aghavani, A., Arcanzi, T.T., Baxter, E., Blazef, R.G., Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B., Lomoton, M.A., Mazda, P., Mok, M.S., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacle, J.M., Park, S., Pfeiffer, B., PUNCH, D., Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R., Zieran, L.L. and Kimmel, B.

TITLE Direct Submission
JOURNAL Submitted (28-FEB-1998) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720. US
and KAMMER, B.

COMMENT

Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720

For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://fruitfly.berkeley.edu/sequence-archive.html>) or send email to drosophila@genome.tbl.gov.
Library location: 136.81.

library location: 136_01.
This pl was assembled from the following subclones: 1_b4, 2_b6, 2_d8, 1_h5 (AC003251), 1_c11, 2_e1, 1_f2, 1_g4 (AC003249), 1_g6 (AC003250), 2_c9 (AC003252), 1_f12 (AC003248), 1_h9, 2_e9, 1_a5, 2_e3, 1_e12 (AC003247), 2_h1, 2_b12, 1_a12 (AC003761), 1_g3, 1_g12 (AC003249), 2_e7, 1_c8.

```

FEATURES
source
1. .64674
Location/Qualifiers
/organism="Drosophila melanogaster"
/db xref="taxon:7227"

```

25

```
/chromosome="2L"
/map="26F1-26F4"
/clones="P1 DS07773 (D154)"
BASE COUNT 18746 a 13758 c 13808 g 18362 t
ORIGIN

Query Match          6.5%; Score 90.8; DB 3; Length 64674;
Best Local Similarity 45.48; Pred. No. 1.5e-12;
Matches 556; Conservative 0; Mismatches 632; Indels 36; Gaps 5;

Qy 47 catttccaaagagacaagtacaaatttctctactgtgctgctgctcattcttctgttgattcggag 106
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 59747 CAGCGCCAAAAGATAAATTCCTTATTCGTGTGTTTATATCTGCTGCTGACGCGGTGGSCA 59688

Qy 107 ttctctgcaatgaaatgttctattactatcgccccc-----tgagattatgtgaatt 160
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 59687 CGCTGATGCCATGAACATGTTTCATCAGCGCCAAAGTCCCTATTTGAGGAGCACTCAAGTTTCG 59628

Qy 161 attggttcaaacccggatggcggtggagacatggtatttcgaagaattcatggatctttga 220
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 59627 GACCAAAACACACCGTGGCCAGGAGTGAAGTACCGGACCCCAATTCATGCACAAATATGG 59568

Qy 221 cgattggttcacaacttccaaacgcaagcattaatgttttcaacgtgttctctcatttg 280
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 59567 GCTTTGCTCGCAGATTCCTAACCTGGTCTTCAACTGGCTGAACATCTTTGTCAACTTTG 59508

Qy 281 ctgttccctgaatctacgcgtctttgtctcgggtttgttcaacatcgtaaacctgacaa 340
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 59507 GCGGGCGGCTGACCCACCGAATCGTCTACAGCATCATCTTGGAAATGTCATCTTGTCTGG 59448

Qy 341 tcaattctcatctctgtcattgttctgagcccaatgaagattccatgttcctgttttct 400
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 59447 TTACCATTTATTTGGCCATGCTCG-----ACTCTCGCAGTGGCGGCGTCTTCTTCT 59394

Qy 401 gggtaactcttggaaatggcgacttcaatcaatttttagcaatgggctatatgaaactcgg 460
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 59393 GGACACCATCGTGTGCAATGTTGTTTAAACGTGTGCAAGCGGCATCTACAGAACACCA 59334

Qy 461 ttattggagttgggagattttccgcacactcattggcgctctcttggattggaaca 520
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 59333 TATACGGAATAGTGCATCTGTTGCCATCAAGTACACAGGCGCCGTCGCTCGGCTCCA 59274

Qy 521 acatttgcgagttgtgtataacggttggaaaatcgagagtgacctattttctggaatgatg 580
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 59273 ACATCAGCGGTGCTTTCACCCAGCGCATGGCTTGATATGCGGTGAGATCTTCTCGTCCA 59214

Qy 581 agcctaaacttgttgaatctctatttctgcaatctctgttggatcctcttctgtgtgtg 640
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 59213 AG---AGGACTTCGGCCATCTACTTCTGTGACCGGCATCTAGTGTGCTGTCTGCT 59157

Qy 641 caattgcaacttttctatcaaaagcaagatttctaccactatcaccatcaaaaaggaa 700
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 59156 TCGACACGACTACTTTCGGCTGCGGTGAACAGTCTTCCGGCAGTACGAGACCATTAGCC 59097

Qy 701 tggaaattcgcgaaagcggaacccgacagacgctctccatccattcttttggaccacat 760
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 59096 GGAGCAGCGAAGAAATCTGATTCTAAGGCCACAGCTGACGTCGCTTACTGGCAGATCT 59037

Qy 761 tcaaaaacttattgggcaactcttcaattgtttggttctgtctgttgccttactctcaca 820
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 59036 TCAAGAAAGACGAGCGCCCAACTATTTCACATCTTCTCAGCTTTTTCGTCAGTTGTGGG 58977

Qy 821 tcttctctgttatgatgaccgttaccactctggagagattcggcttctctcaaaaatta 880
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 58976 TTTTCCCGCGGATTCAGTGAAGTGCACCGTCCGACCCCAATTTCTGTA----- 58925

Qy 881 tgttgaaaacgatgaaatctacacttttgcacaaagtttctctcgtcttctcaattgttcg 940
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 58926 -----GTAGGCCCGGACTACTTTACACTGGTCACTTGCTTCGCGACCTTTAACGTCCTCG 58872

Qy 941 ctgctgattggtatccatagttgtcttccaagattcactggtccgacaccccggttacctcaaat 1000
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
```

```
Db 58871 CAATGCTGGGAGTTTAAACCACATCGTGGTGCAATGTCGCGGTCCGAGTTCCTCTGGG 58812
Qy 1001 ttgccaataatctcgctgctcttttctcattctctctctctctctctctcgaacta-----tc 1054
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 58811 TGCCAGTGGTCTTGGCTGCTGGGTTTCATCCCTCTCTGCTGTCATGTCGAACATACGTTCCGC 58752
Qy 1055 gtgtccagacgcgtgcttctctctgttttcttctgtgactctactgacatttttctgtgattggtg 1114
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 58751 CGGACTCGGTGGCTCATTTGGCCGTATTTCATCGAGAACGATTGGGTGCTACTGGGGCATCG 58692
Qy 1115 gaattgcaatgtcttttccatggtgataccctcagcgtctggaatgggatacactccaa 1174
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 58691 GCATGCGAATGCCCTACAGCTCCGCGTATTCAGCTCTCTGCGCATGATGATGACGCCCGC 58632
Qy 1175 acgtcgtaccatctcactactcaagatttgcgctcgaactttccgttttgcactcttatgg 1234
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 58631 ACAGGTCACACCAAAATACCAAGACCACTGCCGGAATGTACGACGCTGCCATGCTGATCA 58572
Qy 1235 ttgacctctcacgcgtggtgcctgt 1258
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
Db 58571 CGGGCATCTTCTCCGAGTGTGT 58548

RESULT 4
AC092230 174376 bp DNA INV 30-JUN-2001
LOCUS Drosophila melanogaster, chromosome 2L, region 26F-27X, BAC clone
DEFINITION BACR25K24, complete sequence.
ACCESSION AC092230.1 GI:14578111
VERSION AC092230
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 174376)
AUTHORS Celisner,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Bazon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Scheeler,F.,
Shouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Unpublished
Sequencing of Drosophila chromosome 2L, region 26F-27X
2 (bases 1 to 174376)
Celisner,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Bazon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Shouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Direct Submission
Submitted (30-JUN-2001) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
```

Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bdgpe@fruitfly.berkeley.edu.

FEATURES	source
Location/Qualifiers	
1. .174376	
/organism="Drosophila melanogaster"	
/strain="y; cn bw sp"	
/db_xref="taxon:7227"	
/chromosome="2L"	
/map="26F-27X"	
/clone="BACR25K24 (D1294)"	
/clone_lib="rPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBACE3.6)"	
48681 a 38138 c 38499 g 49058 t	
BASE COUNT	
ORIGIN	

[illegible]

Qy	761	tcaacaadgttattgaggaaactctccaagttttggttcgttgcgttactctaccaa	820
Db	40011	TCAAGAAGACGGCCCAACTATTCAACATCTTCTCAGTTTTCGTCACAGTTGTGG	40070
Qy	821	tctccctgttatgatgaccgtttaccactcgtggagatcccgcttccatacaaaaatta	880
Db	40071	TTTTTCCCGCGGANTTAGTGAACGTGCACCGCTCGACCACCAAAATTCGTA	40120
Qy	881	tgtctgaaaaacgatgaatactcacactttgtccacaagtcttcctcatatttgttcg	940
Db	40121	----GTAGGCCCGGACTACTTTACACTGGTCACCTTGCTTCGCGACCTTTAACGTCCTCG	40175
Qy	941	ctgcgattgatcatagttgtcttccaaagatctcaactgcccagacacccogtttacctcaaat	1000
Db	40176	CATGCTGGGACGTTTAAACACATCTGTGGGTGCAATGGCGGGTCCGAGGTTCTCTTG	40235
Qy	1001	ttgccataactcttgctgctcttttcatccattcttcttcttgcgaacta-----tc	1054
Db	40236	TGCCAGTGGCTTTCGGCTCTGGCGTTTCATCCCTTGTTCATGTGCAACTACGTTCGCG	40295
Qy	1055	gttccagaacgctgcttattacctgttttcttttagcttactagacaatttttgugattgg	1114
Db	40296	CGGACTCGGTGGCGCTCATTTGGCGGTATTTCATCGAAGACGATTTGGGTGTACTGGGSCATCG	40355
Qy	1115	gaattgccatgtcttttccatagatacctcagcgtcttgcaatggatcacactccaa	1174
Db	40356	GCAFCGGAATTTGGCTTACAGTCCGGGTATCTCAGCTCTCTGGGCATGATGTACGCCCGCGC	40415
Qy	1175	acgtctgycatctcaactacaagaattggcgcgtcagcttcttcggtttgcactctta	1234
Db	40416	AGACGGTGCACACAAAATACCAGACACTGGCCGAATGTACGCAGCTCCCATGCTGTATCA	40475
Qy	1235	tggccttctccacggtggcctgt	1258
Db	40476	CGGCAATCTTCTCCGGAGTGCTGT	40499
RESULT	5		
AE003614	259464 bp	DNA	INV 04-OCT-2000
LOCUS	Drosophila melanogaster genomic scaffold 142000013386055 section 7		
DEFINITION	of 63, complete sequence.		
ACCESSION	AE003614 AE002690		
VERSION	AE003614.2 GI:10728613		
KEYWORDS	HTG		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;		
AUTHORS	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 259464) Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amann,T.P., G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Bayler,F.G., Helt,G., Nelson,C.R., Gabor		

RESULT	5
AE003614	
LOCUS	DNA INV 04-OCT-2000
DEFINITION	Drosophila melanogaster genomic scaffold 142000013386055 section 7 of 63, complete sequence.
ACCESSION	AE003614 AE002690
VERSION	AE003614.2 GI:10728613
KEYWORDS	HTG:
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 259464)
AUTHORS	Adams,M.D., Celisner,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Vandeil,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor Miklos,G.L., Abril,J.F., Aghayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Bernan,B.P., Bhandari,D., Bolshakov,S., Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burris,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,S., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrieria,S., Fleischmann,W., Foster,C., Gabrielian,A.E., Garg,N.S., Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Thewissen,C., Talati,M., Kalush,F., Karpen,C.H., Ke-Z.

Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lal, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R., Scheeler, F., Shen, H., Shue, B.C., Siden-Klamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Williams, S.M., Woodage, T., Weinstein, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.

The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
20196006

2 (bases 1 to 259464)
Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
Direct Submission
Submitted (21-Mar-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7297129.
Location/Qualifiers
1. 259464
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2L"
join(<538..1127,1190..1366,1421..>1745)
/gene="CG9550"
/product="CT27016"
/db_xref="FLYBASE:FBan0009550"
/db_xref="FLYBASE:FBgn0031826"
/note="CG9550"
/map="26D7-26D9"
/db_xref="FLYBASE:FBan0009550"
/db_xref="FLYBASE:FBgn0031826"
join(538..1127,1190..1366,1421..1745)
/gene="CG9550"
/note="CG9550 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0009550"
/db_xref="FLYBASE:FBgn0031826"
/protein_id="AAF52397.1"
/db_xref="GI:7297130"
/translation="MVSLSRTSKLTACVGVYTYTVLFIFFLLPMLPDPVTLQRSIV
SYSTSLGIDIRNLTLDAGQPIRSMVLTFRSGALTLDDNLAHQPGCVQHYAPLIAVE
SRSTAKEHRLDELAAALYNCYDKSAEMIQWGRSPVFRFYCAQAKICRSYQSEA
CWDEPTNAVCKLYPFINMAYNLRNLASLLERKDLNLMILLVRDPRTGIYRMN
NKWSDERDCQALCGDMVSDYQMETLTQAYPQRFSLIRYEDLFLQPDSEIKQVED
FYGLPLERNTRIQMKHPRSGYFFQSSRWKLSQSAFSEWSEMMVDDIGAVQDVCGQ
AMDLMGYRSIQDHNHFSPEFQIMDKAY"
2069..>3227
/gene="CG9551"
/product="CT27018"
/db_xref="FLYBASE:FBan0009551"
/db_xref="FLYBASE:FBgn0031827"
/evidence=not_experimental
<2069..>3227
/gene="CG9551"
/map="26D7-26D9"
/db_xref="FLYBASE:FBan0009551"
/db_xref="FLYBASE:FBgn0031827"
/evidence=not_experimental
2280..3227
/gene="CG9551"
/note="CG9551 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0009551"

/db_xref="FLYBASE:FBan0031827"
/evidence=not_experimental
/protein_id="AAF52398.1"
/db_xref="GI:7297131"
/translation="MSTGIGSSSSSTEQOHLSEADLEACSLILGPQDPQSVNVTGRTR
AKQLSLNGKRRRRSRRANLIGCVSSLCILLIATTHRPLTTPFNQSGARDQO
ASVGPGSNLSARLAFYDRFOQOQQSQATVYNLTATVDVLSAQRSLIAEME
NFEPGRGAELTDMTPETNGTPVRSVVTSWRSSTFLGDILNSIPGNFYHPEPLD
FGIKQIPDDQDELAVONKLLNCDVADMIDYLNFGKTHTYLFEHNTRLMDVCREFP
RGCWRPAFLTRFCBLFPIQSMKTYRLRLAOKLLEOR"
join(<27409..27443,28126..28308,28377..28578,
28641..>28772)
/gene="CG13765"
/product="CT33249"
/db_xref="FLYBASE:FBan0013765"
/db_xref="FLYBASE:FBgn0031828"
/evidence=not_experimental
<27409..>28772
/gene="CG13765"
/map="26D9-26E1"
/db_xref="FLYBASE:FBan0013765"
/db_xref="FLYBASE:FBgn0031828"
/evidence=not_experimental
join(27409..27443,28126..28308,28377..28578,28641..28772)
/gene="CG13765"
/note="CG13765 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0013765"
/db_xref="FLYBASE:FBgn0031828"
/evidence=not_experimental
/protein_id="AAF52399.1"
/db_xref="GI:7297132"
/translation="WMVQHVVISPIILSSRVILVLRDPRGTMQSRRRRVWCGGNEDC
EDPLVQDLRDDYKTAAYILLKYPSPRTVRYEDLSLSPSEMTQDILQFGLPDPA
VEEFLDHTKYNIGVGSSTYRDSRSPFHMMDLKEIKOIQDTCFADMLGWVRI
ENFNFTTQQTDFPMWMPPT"
complement(join(30215..30949,31026..31270,31333..33087,
35664..35710,38587..38869))
/gene="cli"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/product="CT27024"
/db_xref="FLYBASE:FBan0009554"
/db_xref="FLYBASE:FBgn0000320"
complement(<30215..>38869)
/gene="cli"
/note="CG9554"
/map="26D9-26E1"
/db_xref="FLYBASE:FBan0009554"
/db_xref="FLYBASE:FBgn0000320"
complement(join(30768..30949,31026..31270,31333..33087,
35664..35710,38587..38658))
/gene="cli"
/note="cli gene product; Nucleotide sequence of the Celera
sequence differs from the published sequence for this
transcript"
/codon_start=1
/db_xref="FLYBASE:FBan0009554"
/db_xref="FLYBASE:FBgn0000320"
/protein_id="AAF52400.1"
/db_xref="GI:7297133"
/translation="MVTLPYNYAAPRCGLIDKMIKPKRPKTDHTDTHERNRLCNL
SQOQOQOQOQOQTHQOQOQOQOQSHSTVLASPSAGAGMGVGVGGGSGG
GGVGGVGCSPGLPQSQPLQPTIGSLASLSGHYSNGNANPNVNSSCSLATASSF
AQSGSFTTYQAGTSGGVSGGVGATVNSHWTHDGTGSSAAVKSRSPQV
HASLDNSVAGSNLYGCSSASNLDGGAVAVNSAAAAAAYDGRKHDYIYNSMQQ
YTPPPFYSGYTPAAATAARAKMEPAAAAAYLTPSYAAGNNNSQLYSSPYAG
YNNFGOODYGGYVNEOYGNYSYSPYSPSSSSASHGHPFAASNSLSEPTD
THSTTPVHTHSPSLPIPSPTCSGIGPLGNVSAASAAAALNSGSGSVCTAGSGG
VATSKTPTGKTGRARRHQPSPTSTRSTASDTGNSEAVKPERVVDLDTLFIH
TLLSGYANRYTKDHSUMTIAFRMEVMFNADTHFFNEIECDQVHIDVSDDN
QDLSAYNATDGFHTNTPPGAPPNCLPTVGRGVDMWRKLAFRYKIKDIYNSYRG
NVGTILGPGKEAWLIQRISIEVATDNWATLALKLSMISQRENCVNVLTSTQLAPA

Qy	761	tcaacaactgttatg	ggcaactcttccaatg	ttttggtctgtcttctg	cggttactcttcaaaa	820
Db	66312	TCAAGAAAGACAG	CGCCCAACTATTTC	CAACTTTCCTCAG	CTTTTCGTACGTTG	66371
Qy	821	tcttccctgttatg	atgacggttaccactg	tgtggagatctcggct	tctctcaaaaaatta	880
Db	66372	TTTTCCGGCGCAT	TTCAGTCGAACGT	CGACCGCTCCG	ACCCAAATTCCTGA	66421
Qy	881	tgtctgaaaaacg	atgaaactctacact	tttgtctccaaa	gtttctctgtcttca	940
Db	66422	-----GTAGG	CCCGACACTTTAC	ACTTGGTCACACT	TGCTTGGCGACCTT	66476
Qy	941	ctgcgattggtac	catagttgcttcca	agattcactggtccg	acaccccggttacct	1000
Db	66477	CAATGCTGGG	CAGTTTACCACAT	CGTGGGTGCAAT	GATGCGCGGGTCCG	66536
Qy	1001	ttagcataaatct	tgtggtctctttt	ctattccattctt	cttcttctgcacata	1054
Db	66537	TGCCAGTGGTCT	GTGCTCGGGGTTC	ATCCCGCTGTGTT	GTGATGTGCAACT	66596

[illegible]

It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone F44D12 is at 1 in this sequence. The true right end of clone F44D12 is at 6268 in sequence 268299.

The true left end of clone T04B2 is at 31750 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence 268882.

The end of this sequence (31750..31853) overlaps with the start of sequence 268299.

For a graphical representation of this sequence and its analysis see: <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F44D12>.

FEATURES

source

Location/Qualifiers

1..31853

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="IV"

/clone="F44D12"

join(3277..3413,3774..3951,3994..4239,4318..4603,

4647..4894,4945..5180,5225..5395,5449..5759,5805..6053,

6100..6802,6846..7020,7211..7306,7357..7425)

/gene="F44D12.1"

Join(3277..3413,3774..3951,3994..4239,4318..4603,

4647..4894,4945..5180,5225..5395,5449..5759,5805..6053,

6100..6802,6846..7020,7211..7306,7357..7425)

/gene="F44D12.1"

/note="Similarity to Mouse Tight Junction protein ZO-1

(SW:Z01_MOUSE), contains similarity to Pfam domain:

PF00595 (PDZ domain (Also known as DHR or GLGF).),

Score=35.0, E-value=5.5e-07, N=2

cDNA EST yk16d12.5 comes from this gene

cDNA EST yk45f4.5 comes from this gene

cDNA EST yk35f8.5 comes from this gene

cDNA EST yk129c3.3 comes from this gene

cDNA EST yk129c3.5 comes from this gene

cDNA EST yk360b12.3 comes from this gene

cDNA EST yk360b12.5 comes from this gene

cDNA EST yk352d5.3 comes from this gene

cDNA EST yk352d5.5 comes from this gene

cDNA EST yk324a9.3 comes from this gene

cDNA EST yk324a9.5 comes from this gene

cDNA EST yk219a4.3 comes from this gene

cDNA EST yk219a4.5 comes from this gene

cDNA EST yk218f7.3 comes from this gene

cDNA EST yk218f7.5 comes from this gene

cDNA EST yk220g3.5 comes from this gene

cDNA EST yk252d9.5 comes from this gene

cDNA EST yk357f4.5 comes from this gene

cDNA EST yk388f10.5 comes from this gene"

/codon_start=1

/protein_id="CAA92607.1"

/db_xref="GI:3871158"

/db_xref="SPTREMBL:Q20398"

/translation="MEKEIREVEAKMEKVRKASERDDIGKKIDFEKIDSLQIQV
HRSFTYKEEANKRLRLQYALQDEYTEMORSIVLEENRDEERDRQKIEK
MKDESEVAALKLENTRLKLVNMEETAQANKEAIAARLTAENGARITFERD
EALRKCQQTGDAIVDINWTHSVQINLPNPNLGLIMGGRTDGSIVGHYIVRQI
AHGSPFDNLKLDHIMVNDISVTDMDERSVGMGLSNCHILHLVRRSNCKISDV
CLPNYGLBELSGVFINSCPEGAASRGLAQGVVHVHMTPTVYDAKHAEMLIK
SREPLVIGILASTRGDNGDKHRQTIFSRWFSRNGSGDKERTVYAKANDRSNDQV
LLRGLSLRMPQASPMSPLYRGSRAPTVSSSIDHTKMLDALDKRFAKSTHSE
ISSSTAAPTALWSPSSVNEKDLVFNAGVVPVVPKSTFESTPTSPVPLIQNS
EQKRTMSQISSHRGWDGRSSNTTSDARPSVMPHTPTSTIMEGRPHRSVAVIS
SHPPVPTIRDDSSMSSNSIRLPTSPFSNOYPNNTSLTGGVPPRHCISKD
GSDFTSITGAMHSYAEGRNKKYHLPRGNHSESLVNRSSQRLVEVPSRSDVLC
GNAIGTIAEKSIGSDLEGLILSDICGNVRNTLEAMNTLSADNSELCTGLVQDG
GDLNRLRLGADGSFFLRVINDRSMENKDELDKCGDVVFDKTMKGTGRWRAWK
VDKRGROEHCARISSTVYQAIRANRYANFPKAYEWEKEKDTKVRKRPVLLFGAVV
EPFLQMLVDESDKTSIVARESLADDEVSAALLDKVLIDSKQNDVYDLVHVSTAH
IMDTAQGLHCLQVQDSADILRAKRCRFFILVIRFKSVKQLKDVNBHIGEKISSK
EAKQLIEKDLKIEKDLGDSVTLVVPSSHNNVFMTHAVLQLKLLIEDSQKLVWVQRK

gene

CDS

VDEE"
complement(join(7556..7673,7717..7832,8096..8296,
8340..8510,8562..8718,8944..9111,9541..9638))
/gene="F44D12.2"
complement(join(7556..7673,7717..7832,8096..8296,
8340..8510,8562..8718,8944..9111,9541..9638))
/gene="F44D12.2"
/note="cDNA EST yk169e12.3 comes from this gene
cDNA EST yk367d4.3 comes from this gene
cDNA EST yk367d4.5 comes from this gene
cDNA EST yk553h9.3 comes from this gene
cDNA EST yk553h9.5 comes from this gene"
/codon_start=1
/protein_id="CAA92598.2"
/db_xref="GI:5824507"
/db_xref="SPTREMBL:Q20392"
/translation="MOTILQILIVLFTLFHEVNLVTCRCRFGNECNFSRSTRCGD
FCYHAQFYGNEMMPTIQKGIVGEVAGECRVNHHGVMFCFCSADYCNANTSLAD
TATTLPTQTCQPEKVNMPKRWTKPCAANYCTFTEAKTQTEVNSANTWTSCNKK
ENEFDFPPTLVFNFPYGTVCVNLNYSRGQPDTHACYSGLDITLFDVTTATTECHVD
YFNRLPYVKSGSSCIGQFCFISATSRGVEGCVNSVTVEGATPLKIGVTRAYTGLE
QNICQSYCNADLKSLELWPPPELYLYRNISNLREFNVFIDSARSSSIFLAIPVF
LNYLCUFH"
join(12326..12380,12427..12481,12574..12895)
/gene="F44D12.9"
join(12326..12380,12427..12481,12574..12895)
/gene="F44D12.9"
/codon_start=1
/protein_id="CAA92605.1"
/db_xref="GI:3877156"
/db_xref="SPTREMBL:Q20396"
/translation="MTKHMQQEESPLSVFYQYLHVEYNICRKGQTTLKMEELA
ENSSVTGNSQIGLVTPRDTWGRKESPTDRWLYVLIITMHGMLMSNMNFTIAPO
YVHDYFNFNTYQDSEPMIIGVTSQIPNVGIMLTIVMV"
complement(15660..15989)
/gene="F44D12.3"
/gene="F44D12.3"
/note="contains similarity to Pfam domain: PF00635 (MSP
(Major sperm protein) domain), Score=111.7,
E-value=4.7e-30, N=1
cDNA EST yk57e8.3 comes from this gene; cDNA EST yk102a1.3
comes from this gene
cDNA EST yk57e8.5 comes from this gene; cDNA EST yk102a1.5
comes from this gene
cDNA EST yk122a5.5 comes from this gene; cDNA EST
yk149d3.5 comes from this gene
cDNA EST yk151a2.5 comes from this gene; cDNA EST
yk164g1.5 comes from this gene
cDNA EST yk170h1.5 comes from this gene; cDNA EST
yk179d6.5 comes from this gene
cDNA EST yk189f1.5 comes from this gene; cDNA EST
yk149d3.3 comes from this gene
cDNA EST yk151a2.3 comes from this gene; cDNA EST
yk164g1.3 comes from this gene
cDNA EST yk170h1.3 comes from this gene; cDNA EST
yk179d6.3 comes from this gene
cDNA EST yk189f1.3 comes from this gene; cDNA EST
yk19a10.3 comes from this gene
cDNA EST yk92f10.3 comes from this gene; cDNA EST
yk15a11.5 comes from this gene
cDNA EST yk164f1.5 comes from this gene; cDNA EST
yk378f3.3 comes from this gene
cDNA EST yk378f3.5 comes from this gene; cDNA EST
yk431f3.3 comes from this gene
cDNA EST yk431f3.5 comes from this gene; cDNA EST
yk365g9.3 comes from this gene
cDNA EST yk365g9.5 comes from this gene; cDNA EST
yk311h6.3 comes from this gene
cDNA EST yk311h6.5 comes from this gene; cDNA EST
yk299b2.3 comes from this gene
cDNA EST yk299b2.5 comes from this gene; cDNA EST
yk309f7.3 comes from this gene

gene

CDS

cdna EST yk309f7.5 comes from this gene; cdna EST yk303a11.3 comes from this gene
cdna EST yk303a11.5 comes from this gene; cdna EST yk302c7.3 comes from this gene
cdna EST yk302c7.5 comes from this gene; cdna EST yk278d6.3 comes from this gene
cdna EST yk278d6.5 comes from this gene; cdna EST yk264e2.3 comes from this gene
cdna EST yk264e2.5 comes from this gene; cdna EST yk262a3.3 comes from this gene
cdna EST yk262a3.5 comes from this gene; cdna EST yk233c12.3 comes from this gene
cdna EST yk233c12.5 comes from this gene; cdna EST yk232g4.3 comes from this gene
cdna EST yk232g4.5 comes from this gene; cdna EST yk232f9.3 comes from this gene
cdna EST yk232f9.5 comes from this gene; cdna EST yk232b6.3 comes from this gene
cdna EST yk232b6.5 comes from this gene; cdna EST yk232a1.3 comes from this gene
cdna EST yk232a1.5 comes from this gene; cdna EST

Query Match 6.18; Score 84.8; DB 3; Length 31853;
Best Local Similarity 57.08; Pred. No. 4.9e-11;
Matches 155; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

Qy 283 ggtccctgatctaccgcgtcttctgctccggtgttgcctcaacatcgctcaacctgacaatc 342
Db 12947 GGTTCATGATGCTTCGAGTTGTGGTCTCTCATTTGCACTGCATTTGATGGGCTC 13006
Qy 343 attctatcctctgcatgtctgtctgagccactgaagattccatgctctggtttctctgg 402
Db 13007 ATTGTTATTCTTGGCATTCTTTGTAACACCATCTCCAGACAGTGTACATGGTTTATATT 13066
Qy 403 gtaactctggaatggcgacttcaatcaatttttagcaatggcctatagaaactcggtt 462
Db 13067 GTTACTTAATCATCATTAATGCGTATGATGATTTGGCCAAATGGAATTTATCAAAATCTCTGA 13126
Qy 463 tatggagtggtggtgagcttttccgcacacctacatggtgcgtctcttctgattggaaacac 522
Db 13127 TACGAAATCGTCGCGGATTTTCGCGATAAATACATTAACTCTTGGTTATTGGAAACAAT 13186
Qy 523 attgcgagtgctgataaacggttgtaaaat 554
Db 13187 TTGTGTGAGTGTCTTACTTTCAGTATTCAGTAT 13218

RESULT 7
CEC47A4
LOCUS CEC47A4 40664 bp DNA INV 25-OCT-2000
DEFINITION Caenorhabditis elegans cosmid C47A4, complete sequence.
ACCESSION 282263
VERSION 282263.1 GI:3869220
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 40664)
none.
Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
99069613
The C. elegans Sequencing Consortium.
Erratum: [[published errata appear in Science 1999 Jan
1:283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep
3:283(5433):1493]]
2 (bases 1 to 40664)
Mortimore, B.J.
Direct Submission
Submitted (11-NOV-1996) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,

Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwnematode.wustl.edu
On Nov 13, 1998 this sequence version replaced gi:2528885.
Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
[981006 dl] : Cosmid flipped
IMPORTANT: This sequence is not the entire insert of clone C47A4.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true right end of clone C47A4 is at 40664 in this sequence. The
true left end of clone F52d4 is at 101 in this sequence. The true
right end of clone F56f12 is at 9819 in this sequence. The start of
this sequence (1..104) overlaps with the end of sequence 280218.
The end of this sequence (40564..40664) overlaps with the start of
sequence 282273.
For a graphical representation of this sequence and its analysis
see: - http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
name=C47A4.
Location/Qualifiers
1. 40664
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="C47A4"
/complement(join(3491..3973,4192..4254))
/gene="C47A4.1"
/complement(join(3491..3973,4192..4254))
/gene="C47A4.1"
/codon_start=1
/protein_id="CAB62792.1"
/db_xref="GI:6562350"
/db_xref="SPTREMBL:Q9U3L9"
/translation="MTIKDLKIKENFTVYDPSLTLLMDRNGKLKKNKGVDNDEVIK
QLIIDLVDVTGGYKRESITDYFLAMHTIIFPLTIFRYLKTALWYRPAIOKEEYDDDA
KLILIRYIGVSQMEFQKTYTDEIDDLFERECWLKTNCAWKAEADAEKKAQSG
RYKRYKRMKNAGTISFVEDD"
join(15737..15769,15884..16078,16580..16814,17751..18037,
18821..18923,18970..19055,19306..19502,20056..20176)
/gene="C47A4.2"
join(15737..15769,15884..16078,16580..16814,17751..18037,
18821..18923,18970..19055,19306..19502,20056..20176)
/gene="C47A4.2"
/note="contains similarity to pfam domain: PF01733
(Nucleoside transporter), Score=142.5, E-value=2.4e-39,
N=1"
/codon_start=1
/protein_id="CAB62793.1"
/db_xref="GI:6562351"
/db_xref="SPTREMBL:Q9U3L8"
/translation="MVIRYVESISPVDPVPPPLYTQDQNTEDLASTVDGADSDDEHNELI
PEDKGRIVFWIILNGIGVLLPWNMTITIAPOYVDYVFTVNGTATFYADSFQAMGV
VAQVPLIVAILNVLIRGLYRLVLAFLAFNSLLIILALVIFQPSQDQARNWY
IVSLIIVMANNSGLIQNSFEFGNAADFPAPKSNVAVIGTNCICTGTSVLAVATLAF
STQAEVALIYIGLSLILFVCLVSWFCKRMWQCCICVFLVTFVLSVPTVLAFG
QPGYTFPNDVYAGIAVLENFNFFAAVGNAAATEVTFPPRLILVPCVIRLULFPFEM
FSNYLPHSRMTGVLFTNEWIFFFGNTLLAFTSGYFSSGLMMYTPRVCPPEYSKLAGOV
SALSVLGITAGYSTVTAITAMVNNF"
complement(join(28627..28707,28755..28832,28884..29024,
29070..29267,29338..29466,29527..29850))
/gene="C47A4.3"
complement(join(28627..28707,28755..28832,28884..29024,
29070..29267,29338..29466,29527..29850))
/gene="C47A4.3"
/note="contains similarity to pfam domain: PF00149
(Ser/Thr protein phosphatase), Score=476.0,
E-value=1e-139, N=1

```

cDNA EST yk121b5.3 comes from this gene
cDNA EST yk121b5.5 comes from this gene
cDNA EST yk262b6.3 comes from this gene
cDNA EST yk262b6.5 comes from this gene
/codon_start=1
/protein_id="CAB62794.1"
/db_xref="GI:562352"
/db_xref="SPTREMBL:Q9U317"
/translation="MONNVDSIIIDVLSAHSHEKPLCKVITERVVLKLLDLALGVFK
AOKPWEVNAPIKKGDIHGQPDLLRFLHRGWPPTANYLFLGQYVDRGRFSIETIV
LLLAYKYPGCFNFFLLRGNHCEFNKTYGYECCQKRYQSVRYAAQDFVNLPLT
GLIATKILCGHGGSLPLMKFTLDTLRKIERPTGEGKGLVADLLWADPISGLSGFMN
NORGACGFGDSVLNLCSEFQDLVRAHQVQDGYFFAGRKLVTITFSAPHYCCQF
DNCAAFMSCDEKLOCSEFILRPTTGRLEIREKPLKDET"
33294..33521
/gene="C47A4.5"
33294..33521
/gene="C47A4.5"
/codon_start=1
/protein_id="CAB62795.1"
/db_xref="GI:562353"
/db_xref="SPTREMBL:Q9U316"
/translation="MADKSARVPVDAIGNHNKNTDLVDIDDELFCCKPKPKTAEPKPA
PPPPAPAPAAPSTGKYQYKKSSTYKSYAK"
BASE COUNT 13436 a 6918 c 6544 g 13766 t
ORIGIN

Query Match 5.1%; Score 71.2; DB 3; Length 40664;
Best Local Similarity 53.6%; Pred. No. 1.7e-07;
Matches 148; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 390 cgggttttttcggtaaccttgggaatgcgacttaactcaatttttagcaatgggctata 449
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17758 CTGGTTCACATTGCTCTATTATCATAGTAATGCGAATCAATGATCCACGGTCTCTA 17817

QY 450 tgaataactcggttatggaggttggcgatttttcgcgcacactacattggcgctctctt 509
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17818 CCAAAACTCATTTTTCGGAATGGCTGCAGATTTCCTCCGCAAAATACATCGAACGCTGTGG 17877

QY 510 gattggaacaacatttcggatttcctgatacaggttcttggaataacggagtgacctattt 569
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17878 TATTGGAACAAATATTGTTGGAACATTCTACTAGTCTTGGCAATGTGCGAATTTGGC 17937

QY 570 tctgaatgatgcctaaacttctgcaatcgtctctatttcgcgcacactatcgttggatcct 629
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17938 ATTCACACACAGCTGAACATGTGCACTTATTATTATTGGAATATCACTCTTATTCT 17997

QY 630 tctggtgtgcaattgcacttttctttatcaca 665
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17998 GTTGTGTTCTGTCTCATGTGTTGTTCTGCAAGAA 18033

RESULT 8
CEZK809/c 25560 bp DNA INV 20-JUN-2001
LOCUS
DEFINITION Caenorhabditis elegans cosmid zK809, complete sequence.
ACCESSION Z68303
VERSION Z68303.1 GI:1130662
KEYWORDS HTG; Hydrophobic nucleolar protein like; Peroxisome assembly factor
like; Protein-tyrosine phosphatase.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 25560)
none.
REFERENCE Genome sequence of the nematode C. elegans: a platform for
AUTHORS investigating biology. The C. elegans Sequencing Consortium
TITLE Science 282 (5396), 2012-2018 (1998)
JOURNAL 99069613
MEDLINE
REMARK The C.elegans Sequencing Consortium.
REFERENCE 2 (bases 1 to 25560)

```

AUTHORS
TITLE
JOURNAL

COMMENT

Dobson, R.
Direct Submission
Submitted (20-DEC-1995) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rw@nematoe.wustl.edu
Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone zK809.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone zK809 is at 1 in this sequence. The true
right end of clone zK809 is at 15121 in
sequence Z68302.
The true left end of clone zK792 is at 25461 in this sequence. The
start of this sequence (1..113) overlaps with the end of sequence
Z73427.
The end of this sequence (25461..25560) overlaps with the start of
sequence Z68302.
For a graphical representation of this sequence and its analysis
see: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=zK809)
name=zK809.

Location/Qualifiers
1..25560
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="zK809"
join(1372..1376,1431..2001,2054..2551,2599..2803,
2847..2854)
/gene="zK809.1"
join(1372..1376,1431..2001,2054..2551,2599..2803,
2847..2854)
/gene="zK809.1"
/note="contains smilarity to Pfam domain: PF00102
(protein-tyrosine phosphatase), Score=35.2,
E-value=2.3e-13, N=1
cDNA EST yk365g6.3 comes from this gene"
/codon_start=1
/protein_id="CAA92641.2"
/db_xref="GI:5824906"
/db_xref="SPTREMBL:Q23602"
/translation="MPNKGVGKKSPSTSKSSNNKKRRREPESRYEEKTAVHGGTKTYE
EKTVVHGGTKNVVAKVPKASVIVDKKASTTETTPITVKKSMLSISYADRTVTE
EXEKPKKKEKESGEDASIKFELRENOSEGVRDTEASCNPKCKWQSHSRN
RHTDYKYNHRIIVOMCKDDYINGSKINVPHTPIYLVLPKMSVDVADVEFRVW
FHEQCQTVHIIARPELTNPAIDKLCFQSGAWLYANGFFVTRKVEKKNADMPV
VELLPCCSNAYMVCVYLHTYWKPLQFPRFGAIIAAHQIAKNKNGNSPTVIASNG
SGRNAALLTAVVEDQLTGRGPKIETVIRTIQEQPSQSDVSIQVLSYMATNWLK
TKLPSCHEVVSIIKKFNEKATRSIPLGSMKTPTEEL"
join(3252..3612,3660..3835,4567..4606,5066..5406)
/gene="zK809.7"
join(3252..3612,3660..3835,4567..4606,5066..5406)
/gene="zK809.7"
/note="Similarity to Rat peroxisome assembly factor-1
(SW:PAFL1-RAT), contains similarity to Pfam domain:
(zinc finger, C3HC4 type (RING finger)), Score=16.2,
E-value=0.00035, N=1
cDNA EST yk360all.5 comes from this gene"
/codon_start=1
/protein_id="CAA92640.1"
/db_xref="GI:3881798"
/db_xref="SPTREMBL:Q23601"
/translation="MIHFRYHOKRNILPNTLHCTLPHYHYFSVFLSDTSFYLIIF
SFVSTFLYFKMACRQDLDAHVLDSEIRSIADQSDVDDVNMPLWVARIFEKLR
PELKVTLEAVLWTRFSRGASGSEQMDIAYAGYTPKLTITSHFMVEVIVPILTRMSE

LSGRIDISRMYSKIEAIYELGSLHFLYFLRKSASINYDTONRELMWHAFRDVILLTYP
FIEKARIVIKOKLRKFTSSIGSDIECVKDCPSVIMVGOKCHVACYCTIATS
RNTWCPUCSNAEMPELAKLNPLIE"
complement(join(10131..10337,10438..10717,10774..10870,
10916..11035,11085..11127))
/gene="ZK809.2"
CDS
complement(join(10131..10337,10438..10717,10774..10870,
10916..11035,11085..11127))
/gene="ZK809.2"
/note="contains similarity to Pfam domain: PF01553
(Acyltransferase), Score=144.3, E-value=7e-40, N=1
cdna EST yk576b1.3 comes from this gene"
cdna EST yk576b1.5 comes from this gene"
/codon_start=1
/protein_id="CAA92638.1"
/db_xref="GI:3881796"
/db_xref="SPTREMBL:Q23598"
/translation="MSLVTSVKLMFLGGSKNLICHNKETFKVILENPPLITVSNH
RSNIDPLMWCILKREFWYKDRNRYTLAAHICFTKQFHTTMSLGRVPCVGRGEG
VYQKMDFCVMDLNKRWVHIFPEGKVTCTLESPFLRFKWKIGRLVMDAKTDPVILPVW
CKMEKVMPTQPPYYPKFGNTVTVHIGEPFLSDLKTVLSKSLTTEQMRKIITDEVO
TRNVLGEKVGDLPGKSSLEILRKNPIEV"
join(11625..11731,11785..11887,11940..12056,12122..12232,
12286..12419,12469..12529)
/gene="ZK809.3"
join(11625..11731,11785..11887,11940..12056,12122..12232,
12286..12419,12469..12529)
/gene="ZK809.3"
/note="predicted using Genefinder
cdna EST yk78b3.3 comes from this gene; cdna EST yk78g8.3
comes from this gene
cdna EST yk105h6.3 comes from this gene; cdna EST yk78b3.5
comes from this gene
cdna EST yk78g8.5 comes from this gene; cdna EST yk105h6.5
comes from this gene
cdna EST yk636e8.3 comes from this gene; cdna EST
yk636e8.5 comes from this gene
cdna EST yk437a12.3 comes from this gene; cdna EST
yk437a12.5 comes from this gene
cdna EST yk459f11.3 comes from this gene; cdna EST
yk459f11.5 comes from this gene
cdna EST yk456e1.3 comes from this gene; cdna EST
yk456e1.5 comes from this gene
cdna EST yk435c7.3 comes from this gene; cdna EST
yk435c7.5 comes from this gene
cdna EST yk435b11.3 comes from this gene; cdna EST
yk435b11.5 comes from this gene
cdna EST yk417d5.3 comes from this gene; cdna EST
yk417d5.5 comes from this gene
cdna EST yk391b6.3 comes from this gene; cdna EST
yk391b6.5 comes from this gene
cdna EST yk197b7.3 comes from this gene; cdna EST
yk197b7.5 comes from this gene
cdna EST yk199b5.3 comes from this gene; cdna EST
yk199b5.5 comes from this gene
cdna EST yk538d11.3 comes from this gene; cdna EST
yk538d11.5 comes from this gene
cdna EST yk561e2.3 comes from this gene
cdna EST yk567a11.3 comes from this gene; cdna EST
yk567a11.5 comes from this gene
cdna EST yk586g8.3 comes from this gene; cdna EST
yk586g8.5 comes from this gene
cdna EST yk616f10.3 comes from this gene; cdna EST
yk616f10.5 comes from this gene
cdna EST yk620g3.3 comes from this gene; cdna EST
yk620g3.5 comes from this gene
cdna EST yk650c6.3 comes from this gene; cdna EST
yk650c6.5 comes from this gene
cdna EST yk665g12.3 comes from this gene; cdna EST
yk665g12.5 comes from this gene
cdna EST yk538d11.5 comes from this gene; cdna EST
yk538d11.5 comes from this gene
cdna EST yk567a11.5 comes from this gene
cdna EST yk586g8.5 comes from this gene; cdna EST
yk586g8.5 comes from this gene
cdna EST yk616f10.5 comes from this gene; cdna EST
yk616h5.5 comes from this gene

cdna EST yk620g3.5 comes from this gene; cdna EST
yk623f9.5 comes from this gene
cdna EST yk650c6.5 comes from this gene; cdna EST
yk655c5.5 comes from this gene
cdna EST yk665g12.5 comes from this gene"
/codon_start=1
/protein_id="CAA92637.1"
/db_xref="GI:3881795"
/db_xref="SPTREMBL:Q23597"
/translation="MGNQYKPKERTGNSPYTRSGVPPALVKTGDKPLHRVNDMSLE
LHADERVRAAGLSPAERWRKRWQDHLHADEPVVDVHRQLNPIRTAYRLPWDK
FYHLKPTFGVYGTATYAPKLFMAFVVQYATYKYEVKDWTHRLGLESMPQK
EVIASAPAEIEKPEGLMDKGLANPSSKDDYTFYFNKRTAYLDVGETRPW"
complement(join(12918..13232,13332..13978,14028..14122,
14201..14367,14519..14680))
/gene="ZK809.4"
complement(join(12918..13232,13332..13978,14028..14122,
14201..14367,14519..14680))
/gene="ZK809.4"
/note="Similarity to Human 36k hydrophobic nucleolar
protein (PIR Acc. No. Jc4196), contains similarity to Pfam
domain: PF01733 (Nucleoside transporter), Score=541.4,
E-value=2e-159, N=1
cdna EST yk77f11.3 comes from this gene
cdna EST yk77f11.5 comes from this gene
cdna EST yk373c3.3 comes from this gene
cdna EST yk373c3.5 comes from this gene
cdna EST yk328c5.3 comes from this gene
cdna EST yk328c5.5 comes from this gene"
/codon_start=1
/protein_id="CAA92642.1"
/db_xref="GI:3881800"
/db_xref="SPTREMBL:Q23599"
/translation="MSSAVELQPLNKTKKVEESEPEDKGNLVFYIILLHGICGLMP
WNMLITISDYDFESYKMLANSTIDMTGVYGYPTVYSSNFSQQTASQVPLLNL
LNIFIVVKGGLASRITVGLSIVACVITTMFIYVETSTWLTGFTTITITIIVLNGA

	Query Match	5.1%;	Score 70.8;	DB 3;	Length 25560;
	Best Local Similarity	52.0%;	Pred. No. 2e-07;		
	Matches 159;	Conservative 0;	Mismatches 147;	Indels 0;	Gaps 0;
Qy	978	gcacacaccccggttacctcaaatgtgcatatcttgcgctgctcttttcaattccattttt	1037		
Db	13233	GCCTCGCGCGCAACAAAGCTCTGGATTCCAGTCTACCTCGCTCTCTATATTCATTCTT	13174		
Qy	1038	ctctcttgcacatcgtgTccagacgcgtcttctctcttcttgcagctactga	1097		
Db	13173	CATCTTCGCAACATATCTCCCTGAGACCCGTTCACTTCCCGTCTTTTTCGAATCCACCTG	13114		
Qy	1098	cattttgtgattggtggaattgcatgtctttttcacatggataccctcagcgtcttggc	1157		
Db	13113	GCTTTTCGTCATTATCGCGCGCTCGATGAGCTTTGGAAGTGGATATTCTCGGGACTCGC	13054		
Qy	1158	aatgggatactaccacaaacgtcgtgcaatctcactactcaagattgcccgtcagctttc	1217		
Db	13053	CATGATGTACACCTCGAAGACTGTTGATCCATCGAAGCTCAGGTCGCTGGAATGATGC	12994		
Qy	1218	cggttgcaactcttatggttgcccttctcacccgtggccctgtggccggttgattatgagca	1277		
Db	12993	CGGAATCTTCCTCATCTCTGGAATTGCTCTGGTCTGATCTTACATGGTCATCAAGAT	12934		
Qy	1278	cttcgt 1283			
Db	12933	GGTCGT 12928			

RESULT 9
LOCUS I66494 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS

```

SOURCE      Unknown.
ORGANISM     Unclassified.
REFERENCE    1 (bases 1 to 7218)
AUTHORS      Dorner,F., Scheflinger,F. and Falkner,F.Gunter.
TITLE        Recombinant fowlpox virus
JOURNAL      Patent: US 5670367-A 14 23-SEP-1997;
FEATURES     Location/Qualifiers
             source
             1. 7218
             /organism="unknown"
BASE COUNT   1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match      4.9%; Score 68; DB 6; Length 7218;
Best Local Similarity 4.2%; Pred. No. 9e-07;
Matches 17; Conservative 234; Mismatches 149; Indels 0; Gaps 0:

Qy 907 ttgtccaaagtctctcgtctcaattgttcgctgcgattggatccatagttgcttcc 966
|||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1060 TTGCGATYVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV 1119

Qy 967 aagattcaactggcgcagacccttgctacatcgaattgcataattcgctgctctttc 1026
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1120 YYYVYVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV 1179

Qy 1027 attcattctctctcgaactatcgtgtccagacgcgtgcttctctcttcttctt 1086
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1180 YYYVYVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV 1239

Qy 1087 gagtctactgacattttgtgattggtggaattgccatgtcttttccatggatcacctc 1146
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1240 YYYVYVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV 1299

Qy 1147 agcgtctggcaatgggatactcactccaaacgtcgctgcatctcactcactcaagattgcc 1206
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1300 YYYVYVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV 1359

Qy 1207 gctcagcttctcgttgcactctatggttgccctctcaccggtgctgtgcccgtt 1266
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1360 YYYVYVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV 1419

Qy 1267 gttattgagcacttcgtggacaagcgaagtatcttataaa 1306
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1420 YYYVYVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV 1459

RESULT 10
CEK09A9
LOCUS       CEK09A9          34017 bp    DNA             INV          20-JUN-2001
DEFINITION Caenorhabditis elegans cosmid K09A9, complete sequence.
ACCESSION   279601
VERSION     279601.1 GI:3646501
KEYWORDS    HR6; Aspartyl beta hydroxylase like; NADH-ubiquinol oxidoreductase;
            Protein kinase; RAS-related protein; Ubiquitin carboxyl-terminal
            hydrolase.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditioidea; Rhabditidae; Pelodierinae; Caenorhabditis.
            1 (bases 1 to 34017)
REFERENCE   1
  AUTHORS   none.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
            Science 282 (5396), 2012-2018 (1998)
JOURNAL     99069613
MEDLINE     The C.elegans Sequencing Consortium.
REMARK      2 (bases 1 to 34017)
REFERENCE   2
  AUTHORS   Swinburne,J.
  TITLE     Direct Submission
JOURNAL     Submitted (29-AUG-1996) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:

```

```

jes@sanger.ac.uk or rw@nematode.wustl.edu
On Sep 24, 1998 this sequence version replaced gi:1515146.
Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone K09A9.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone K09A9 is at 35278 in
sequence 279596.
The true right end of clone K09A9 is at 34017 in this sequence. The
true left end of clone K0989 is at 24804 in this sequence. The true
right end of clone C02C6 is at 104 in this sequence. The start of
this sequence (1..104) overlaps with the end of sequence 279596.
The end of this sequence (33914..34017) overlaps with the start of
sequence 279602.
For a graphical representation of this sequence and its analysis
see:- http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
name=K09A9.
FEATURES             Location/Qualifiers
     1..34017
     /organism="Caenorhabditis elegans"
     /db_xref="taxon:6239"
     /chromosome="X"
     /clone="K09A9"
     complement(join(9270..9399,10952..11597,11646..11868,
12089..12263,12317..12450,12502..12642))
     /gene="K09A9.5"
     complement(join(9270..9399,10952..11597,11646..11868,
12089..12263,12317..12450,12502..12642))
     /gene="K09A9.5"
     /note="Predicted using GeneFinder
Similarity to Bovine NADH-ubiquinone oxidoreductase 49 KD
subunit (SW:NUCM.BOVIN), contains similarity to Pfam
domain: PF00346 (Respiratory-chain NADH dehydrogenase, 49
Kd subunit), Score=595.8, E-value=6.4e-215, N=1
cDNA EST EMBL:Z14772 comes from this gene
cDNA EST EMBL:M88875 comes from this gene
cDNA EST EMBL:M89294 comes from this gene
cDNA EST EMBL:T01723 comes from this gene
cDNA EST YK81b6.3 comes from this gene
cDNA EST YK81b6.5 comes from this gene
cDNA EST YK68h5.3 comes from this gene
cDNA EST YK68h5.5 comes from this gene
cDNA EST YK70h1.5 comes from this gene
cDNA EST YK70h1.3 comes from this gene
cDNA EST YK501g3.3 comes from this gene
cDNA EST YK485g3.3 comes from this gene
cDNA EST YK485g3.5 comes from this gene
cDNA EST YK454f8.3 comes from this gene
cDNA EST YK454f8.5 comes from this gene
cDNA EST YK424c6.3 comes from this gene
cDNA EST YK424c6.5 comes from this gene
cDNA EST YK418a5.3 comes from this gene
cDNA EST YK418a5.5 comes from this gene
cDNA EST YK229c10.3 comes from this gene
cDNA EST YK229c10.5 comes from this gene
cDNA EST YK409d1.5 comes from this gene
cDNA EST YK530c9.3 comes from this gene
cDNA EST YK532b10.3 comes from this gene
cDNA EST YK596h9.3 comes from this gene
cDNA EST YK652g6.3 comes from this gene
cDNA EST YK680g8.3 comes from this gene
cDNA EST YK530c9.5 comes from this gene
cDNA EST YK562f5.5 comes from this gene
cDNA EST YK675a3.5 comes from this gene"
/codon_start=1

```

```

/protein_id="CAB01886.1"
/db_xref="GI:3878424"
/db_xref="SWISS-PROT:Q93873"
/translation="MLGRKIAGTCLRANVESVAATSTPATQTRNSHTIWYPAKPER
QFTGTLGLKMSRVSDEIGDKLEKLAISOPVMSDNIYSGKOREKLNEMLIN
FGQHPAAGLGRVRLKLEGEVIAIKAPHLGLHRAATEIKHTKTYQALPYDRLDY
VSMNCQEWASLAEVILGDIIDIPRAKYIRTLMLGELTRIQNHMGITHTHALDVGATMP
FWMFEERKLFFSERVSGARMANVRPGVAVDPLIGLMDIDYDWAIKFERIDE
LEMLTENRTKARTIDIGLVSADALNMGFSGVMVGRSGIKODVRKTEPYDAYADME
EDYPIGTGKDCYDRICRTEEMRQSLNIVHQCCLKPAGBEIKYDDHKVYPPKRAEKE
NMESLIHRKFFTEGFGVPGVATPIEAPGEGFVTLVADGTGKPKRCIRPAPGFAH
LAAIHVCVNSLIADIVAVIGTMDIVFGEVDR"
join(13620..13793,13874..14074,14437..14532,14631..14697,
14741..14841,14889..15017,15106..15351,15397..15632,
15939..16205,16346..16496)
/genes="K09A9.4"
/notes="Similarity to yeast ubiquitin carboxyl-terminal
hydrolase (SW:UBPD_YEAST), contains similarity to Pfam
domain: PF00442 (Ubiquitin carboxyl-terminal hydrolases
family 2). Score=55.1, E-value=4.9e-13, N=1; PF00443
(Ubiquitin carboxyl-terminal hydrolase family 2),
Score=12.8, E-value=0.00029, N=1
cDNA EST EMBL:Z14811 comes from this gene
cDNA EST yk527f6.5 comes from this gene"
/codon_start=1
/protein_id="CAB01885.1"
/db_xref="GI:3878423"
/db_xref="SPTREMBL:Q93875"
/translation="MADSGRDDGKHNVDSDSSQNIERKEEGILDEESSSESGTGD
TKDDLEENNPMPKQMESCKHVKDINPDSLQQLKCLKMCTCHDSVSFSCAH
QDINSLSCKCKNSHLLPHMAVHNHPILYHDKSILCTRCMCTVTPVRFRTHTKF
STSDQETRKNEPLAPRGLLGLYLNFGNTCYMNAVLLQLGHCSPTQYLIELIPGG
WNSCDEIPTKALQMALDURNMYSDPFLYLSPWKIIITQDASEFMRNLDLIDRLDK
TCSEYIRTLIDEMPEMDYAIAMNAPHTKTAISAFQVLENQCHSCGFRSRTI
ENFLDISIPVIGNEFEEMAYAKLDTKPACTSSNIEDQDPGLVHQAFNFKSLE
SCLDRFONSTLODDNOYSCSKCEVLVDATKSTKAHLPEVIIQIKRPHRTWFGSK
VGVRFEPFLSDQDFGWTGSEPALYDLGVFVHGERSLEFGHVYVSQHEQDNQGR
HVRGSETWEKGFVYQORNAKRWLFASGRHVRGSPHYVSFGYMRLE"
join(23324..23300,23815..23981,24388..24482,24569..25081,
25131..25216,25261..25575)
/genes="K09A9.3"
join(23324..23500,23815..23981,24388..24482,24569..25081,
25131..25216,25261..25575)
/genes="K09A9.3"
/notes="predicted using GeneFinder
contains similarity to Pfam domain: PF01733 (Nucleoside
transporter). Score=501.6, E-value=1.9e-147, N=1
cDNA EST CEMSE20F comes from this gene
cDNA EST yk106d4.3 comes from this gene
cDNA EST yk78h6.3 comes from this gene
cDNA EST yk11h7.3 comes from this gene
cDNA EST yk116g9.3 comes from this gene
cDNA EST yk78h6.5 comes from this gene
cDNA EST yk106d4.5 comes from this gene
cDNA EST yk11h7.5 comes from this gene
cDNA EST yk116g9.5 comes from this gene
cDNA EST yk143f7.5 comes from this gene
cDNA EST yk183c7.3 comes from this gene
cDNA EST yk183c7.3 comes from this gene
cDNA EST yk143f7.3 comes from this gene
cDNA EST yk488b8.3 comes from this gene
cDNA EST yk488b8.5 comes from this gene
cDNA EST yk488b8.5 comes from this gene
cDNA EST yk247h5.3 comes from this gene
cDNA EST yk247h5.3 comes from this gene
cDNA EST yk215h7.5 comes from this gene"
/codon_start=1
/protein_id="CAB01882.1"
/db_xref="GI:3878420"
/db_xref="SPTREMBL:Q93871"

```

```

/translation="MSAAVLQPLNEAKKVEDAEVEETPEPEDKGNLVFYIILLHGI
GTLPMNNLITISYDIPESYKTLANSTIDMTGKVTYGDPTVYSSNQSFOTIASOVPN
LLNLNLFIVKGLAGRTVGLSIVAVCVITMIFIVYVETSWLGFETLITII
VLNANGVYONSIFGLASLPFYKTNVAVIIGNLNCGTFTVLLSMSTKAVRNILDRSF
AYFSIALITLFCFISPHILKROFYQYSTRAEQRAKNEEAAADNEGKNYIATER
EAPFOLINLVFVLTISIFPQNMVYKDEKKGTYDFPLPQNYFMDVTFLOPNVFA
FIGSIVAGRWQPNPKLWIPVLRLLYIPFCNLYPETRTLPLVPFESTWLFIIVA
ASMSFGSGYSGLSMPYTSKTVDPKSAQVAGMMAGPFLISGLIVSLFTVMVKPVVTA
"
complement(join(27534..27696,27760..27878,28371..28615,
28762..28867))
/genes="K09A9.2"
complement(join(27534..27696,27760..27878,28371..28615,
28762..28867))
/genes="K09A9.2"
/notes="Similarity to Rat RAS-related protein RAB-14
(SW:RBI4_RAT), contains similarity to Pfam domain: PF00071
(Ras family). Score=319.6, E-value=1.2e-92, N=1
cDNA EST yk115b5.3 comes from this gene
cDNA EST yk115b5.5 comes from this gene
cDNA EST yk479h5.3 comes from this gene
cDNA EST yk479h5.5 comes from this gene
cDNA EST yk293f2.3 comes from this gene
cDNA EST yk293f2.5 comes from this gene
cDNA EST yk197e3.5 comes from this gene
cDNA EST yk529d11.5 comes from this gene
cDNA EST yk554d5.5 comes from this gene
cDNA EST yk567e9.5 comes from this gene
cDNA EST yk571g12.5 comes from this gene
cDNA EST yk663a11.5 comes from this gene"
Query Match 4.7% Score 65.8; DB 3; Length 34017;
Best Local Similarity 51.0%; Pred. No. 4,1e-06;
Matches 182; Conservative 0; Mismatches 172; Indels 3; Gaps 1;
Qy 978 gccgaacacccgttacctcaaatggccatgaattggtgctctcttccattcattcattc 1037
Db 25260 GCCCGCCCGCAACAGCTCTGGATCCCGAGTACCTCGCTCTCTACATCCCATCTT 25319
Qy 1038 ctctctgcaactcgtgtccgacgctgtctctctctctctctctctctctctctctga 1097
Db 25320 CATCTTCGCAACVACCTCCCTGAGACCCGACACACTCCCGGCTCTCTTCGAAATCCACGTG 25379
Qy 1098 cattttgtgattggtggaattgcccattgtcttttccatgataccctcagcctctggc 1157
Db 25380 GCTTTTCATCATCTTGGCGCATGATGAGCTTCGGAAGTGGATACTTCTCAGGACTCTC 25439
Qy 1158 aatgggatacactcccaacgctgctgcccattcaccgctgctgctgctgctgctgctg 1217
Db 25440 CATGATGTACACTTCAAAAGACTGTGCATCCATCAAGGCTCAGGTGGCTGGAATGATGGC 25499
Qy 1218 cgtttgcaclcttaagttggtcctctccacgctgctgctgctgctgctgctgctgctg 1277
Db 25500 TGGATTTTTCCTCATTTTCAGGAATCGTCTCAGGTCTCATTTTTCACATATGGTCATCAAG-- 25557
Qy 1278 ctctggtgacaagccaagatctctataataatttatagcattagatatactgttta 1334
Db 25558 -TTTGTGCTCACTGCCTAAATTTTTCAAAAAACTTCTTAACATTAAGACACCTCATTA 25613
RESULT 11
LOCUS PFMAL3P7
DEFINITION Plasmodium falciparum MAL3P7, complete sequence.
ACCESSION AL034559 AL008975 AL008981 AL008983 AL009015 AL010138
AL010143 AL010146 AL010154 AL010157 AL010160 AL010165 AL010169
AL010187 AL010189 AL010207 AL010208 AL010214 AL021885 AL021888
AL034559.4 GI:8052273
VERSION AL139179 Z98556 Z98557 Z98558
KEYWORDS HTG; 40S Ribosomal protein S3A; acyl transferase;
acylaminoacyl-peptidase; ATP-dependent RNA Helicase; cyclophilin;
elongation factor; F49C12.11-like protein; HesB-like domain
protein; histone H2A variant; kinesin-related protein;

```

N-acetylglucosamine-1-phosphate transferase; protein kinase; R-CG7 repeat; R-FA3 repeat; repl1; rep20; rifin; RNA-binding protein; Stevor; T-complex protein 1 epsilon subunit; telomere; var. malaria parasite P. falciparum.

SOURCE

ORGANISM

Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 253305)
Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C. M., Craig, A., Davies, R. M., Devlin, K., Feltwell, T., Gentles, S., Gwilliam, R., Hamlin, N., Harris, D., Holroyd, S., Hornsby, T., Horrocks, P., Jagels, K., Jassal, B., Kyes, S., McLean, J., Moule, S., Mungall, K., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutter, S., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Whitehead, S., Woodward, J. R., Newbold, C. and Barrell, B. G.

TITLE

The complete nucleotide sequence of chromosome 3 of Plasmodium

JOURNAL

Nature 400 (6744), 532-538 (1999)

MEDLINE

99376085

REFERENCE

2 (bases 1 to 253305)

AUTHORS

Oliver, K., Bowman, S., Harris, D., Lawson, D., Quail, M. and Barrell, B.

JOURNAL

Unpublished

REFERENCE

3 (bases 1 to 253305)

AUTHORS

Lawson, D., Bowman, S. and Barrell, B.

TITLE

Direct Submission

JOURNAL

Submitted (17-DEC-1998) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

COMMENT

On May 14, 2001 this sequence version replaced gi:2982540
gi:2982541 gi:2982544 gi:2894453 gi:2894464 gi:2982551 gi:2982556
gi:2982558 gi:2982563 gi:2982564 gi:2982566 gi:2894491 gi:2982569
gi:2695974 gi:2894502 gi:2673766 gi:2894496 gi:2982577 gi:2894588
gi:2894589 gi:2982538 gi:2982539 gi:2894380 gi:4725992.
For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciparum.

FEATURES

source

location/Qualifiers

1. .253305

/organism="Plasmodium falciparum"

/strain="3D7"

/db_xref="taxon:5833"

/chromosome="3"

/clone="MAL3P7"

458. .4060

/gene="PFC0860w, MAL3P7.1"

458. .4060

/gene="PFC0860w, MAL3P7.1"

/note="PFC0860w (MAL3P7.1), Kinesin-related protein, len:

1200 aa; similarity: to Kinesin-related protein.

D.melanogaster kinesin-like protein 57A (TR:P91945) BLAST

Score: 664, sum p(2) = 1.9e-67; 28% identity in 707 aa

overlap, predicted using hexExon, pfam: match to PR00225

kinesin, kinesin motor domain Score 311.54"

/codon_start=1

/product="kinesin-related protein"

/protein_id="CAB39023.1"

/db_xref="GI:4493964"

/db_xref="SPTREMBL:O97277"

/translation="MRKTNLNDPFLNNYYNKEENPILICSDKERKTKIYSTNH

HTNNILNYDHTLKLICNTENNAKILHSNKSILQKNILMPSYQKQKGTIRETIK

VYPINGEPSTSVENITNGEHFINGQDALKNSLNNDHOHNINNNINNNKILFVL

NNNNNNQVIEENNPTSPITSKNIYDSIYPTNKTNIINSEINNNNNINNNH

NTYINHHNNNNNNPTSSYTSKNNNNLINTLNAYSNVKAIVRIKPIGESENIYS

IFNKKYFIIENKEKECYLILSKKKQSTYVDSVFNVAQEEVFFQAKPLIPHFVK

GINTVFAYGATGSGKTYTMLDDKNQNGIVQLSLELFTIINEKKCRNIKVLMSFLEV

YNTIRDLGLKGNKTLVEQEDVAEVKSNLCELEVNNYEQALLINEGVKNRKMSP

RANKVSRSHALIQIYYNEILDDNMNTISYKALCFVDLAGSERASATSNKGRFKE

GSYINQSLLALANCISLANRNISKVRVYRDSKLTLLKNSLEGNCQVVMIANIP

SRTSFQSNNTAFAPARNIKCATVTDNDKESDIEKTLKKNENLQKVEDTLIGRY

TNLKEFFPIINVLNQLYKKQISCVKLJENISDNNSSMELKODITMYDLQVKMKSDYR

KKVDSLDOQEEKQELNLFDTFLKLNLYVINSKVDNNNNKSLLEMIFFKNEKN

VNENFLYNEKVDKNNVINGVNDENVVDKNNVNGVNDENVVDENVVDENVV

VDNVNVVDENVVDENVVDENVVDENVVDENVVDKVVSNENIVVSEKHKVGLS

AEKGSSEHNKNNKDDIEDNDKDTIKDIHNNNNSSDNDNDEYQSANSPVESDIVKKEK

KKIPINMETKKKRTMNGTKDPIHKTPYDINIVGILNKEDVSNKSNDYNTNKNIEKNKY
EKGEYFPFHNHLMQMSILYINNNVENSPPSRKKKNVAKMLNNGNNNTANFIL
DDDTIKNNNSKISDKNNKNNLNNENNGKINDKSKKKNVNNNNNNNNNNNN
NNNNNSSSSSSGKVDGINILNNSNTNRLHTPSGVYSLNLDIEIKTEINKDKMEKN
DIHLTSDITISKIHHARDLLKENKRKNLENFOENKHEHDEVSLEYVKKKKIKKKM"
4026. .4033

/gene="PFC0860w, MAL3P7.1"

/note="potential splice donor at 3' end of PFC0860w may

indicate splicing"

4057. 4062

/note="potential splice donor at 3' end of PFC0860w may

indicate splicing"

Join(5457. .5577,5689. .5787,5923. .6077,6189. .6743)

/gene="PFC0865w, MAL3P7.2"

Join(5457. .5577,5689. .5787,5923. .6077,6189. .6743)

/gene="PFC0865w, MAL3P7.2"

/note="PFC0865w (MAL3P7.2), RNA-binding protein, len: 309

aa; Similarity eg to C.elegans RNA-binding protein

(TR:Q18318) BLAST Score: 378, sum p(1) = 3.8e-35; 37%

identity in 297 aa overlap, predicted using hexExon, Pfam:

match to PF00076 rrm, RNA recognition motif. (a.k.a. RRM,

RBD, or RNP domain) Score 58.01"

/codon_start=1

/product="putative RNA-binding protein"

/protein_id="CAB39067.1"

/db_xref="GI:4494008"

/db_xref="SPTREMBL:O97318"

/translation="MGLFDKIRNIEKLEAEKLNIGNNDSSWHDOYRDSSYIYIGNLD

NRLPGDIVIFSQYGEIDVNLVDNETGSKGYCFLSYADQDRTILLVDNFNGYKL

LERLPVDHILNRLPKYKADADNEYKQGAEGGIGVYNVSESEIKLSKVFQKIK

NKSEDEKKLLDDEWALNFKSIKKDIISPIGHDEKSRNMGKEEDEDDEDD

DEDDDDDDSDIYKRRHKRKSITTKYDKKHKRKSDDHRDKRRHRNHSRREK

EKKDKSHRRHKHSYDKYSRSRSYSTSSSTDR"

Join(7801. .7928,8084. .8234,8392. .8583)

/gene="PFC0870w, MAL3P7.3"

Join(7801. .7928,8084. .8234,8392. .8583)

/gene="PFC0870w, MAL3P7.3"

/note="PFC0870w (MAL3P7.3), putative elongation factor,

len: 181 aa; Similarity to P. falciparum elongation factor

1-beta, (AAF27524) (276 aa), fasta scores: opt: 366, E(1):

1.1e-15, 53.2% identity in 141 aa overlap, revised:

shortened exon 2, Pfam: match to PF00736 EF1B0, EF-1

guanine nucleotide exchange domain Score 61.10"

/codon_start=1

/product="putative elongation factor"

/protein_id="CAB39068.2"

/db_xref="GI:8052274"

/db_xref="SPTREMBL:O97319"

/translation="MANTYDELYVPLSYIILONEGNTSKIDQANTKKPKKEVINKS

SLIIDIKPYGENTDLDVILKLVNITMEGLTWGKAHKKTTPAFGLFKIQVSCVIVDDL

VNTDELITENLGLDNEQLQKKQMDDEENYDEDDDEITGLVQSAEIIISPNKL"

8235. 8240

/gene="PFC0870w, MAL3P7.3"

/note="revised splice donor sequence for exon 2 of

PFC0870w"

8580. .8585

/note="potential splice donor sequence may indicate

further splicing of PFC0870w"

Join(11431. .20416,20578. .20690)

/gene="PFC0875w, MAL3P7.4"

Join(11431. .20416,20578. .20690)

/gene="PFC0875w, MAL3P7.4"

/note="PFC0875w (MAL3P7.4), hypothetical protein, len:

3032 aa, revised: added 3' exon, possibly spliced at 5'

end"

/codon_start=1

/product="hypothetical protein, PFC0875w"

/protein_id="CAB39024.2"

/db_xref="GI:8052275"

/db_xref="SPTREMBL:O97278"

/translation="MKDPLSYAENNDLNLFLKLYNSEEDCLNKLKYLKSEDFKNVN

NPHLDDLLINHTENKNNLLNLTNINLKSSEFNINDYKLVTFTEFENFKNLKI

NPHLKVVKNTEFYKKNFKIDTKLNKLNHNHYTOYFNDRKKLFFYFNKNNLVETKYC

misc_feature

misc_feature

gene

CDS

gene

CDS

misc_feature

misc_feature

gene

CDS

WE CHINA LINE CS&H/WASING/ADT CONSULTING FOR SEQUENCING PAC CLONES
F6P23, F576, and T13L16, the ESSA group for sequencing clone
F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards

and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES	source	Location/Qualifiers	
		1..110211	
misc_feature	repeat_region	/organism="Arabidopsis thaliana"	
		/cultivar="Columbia"	
		/db_xref="taxon:3702"	
		/chromosome="II"	
		complement(<1..2793)	
misc_feature	repeat_region	/note="Sequence from clone F27L4"	
		complement(356..401)	
		/rpt_family="(TAA)n"	
		complement(join(1549..2058,2731..3012))	
		/gene="At2g23810"	
misc_feature	repeat_region	complement(1549..3012)	
		/gene="At2g23810"	
		/note="F27L4.1"	
		complement(join(1744..2058,2731..3003))	
		/gene="At2g23810"	
misc_feature	repeat_region	/codon_start=1	
		/product="putative senescence-associated protein 5"	
		/protein_id="AAF18611.1"	
		/db_xref="GI:6598449"	
		/translation="MFLILLVFCITVFAFVTVNKGAGEAIEGKYKEYKLGDYSTWLQKRVNKKIRKTSGLSEKVCLEAKFVNVPSFYKEHLTALOGGCKPKSDECGFEVNPWTWNTGTHTPDCOTWDNAKEKLCFDCQSCRAGLLDNVKSARKKVAIYNIVFLVLIIVSYGCCAFRNKRDDSYRHYGYP"	
misc_feature	repeat_region	/note="Sequence from clone T29E15"	
		complement(3567..3677)	
		/rpt_family="(TAAA)n"	
		complement(3570..3641)	
		/rpt_family="(TAAA)n"	
misc_feature	repeat_region	complement(4565..4612)	
		/rpt_family="(CAA)n"	
		6379..6484	
		/rpt_family="(CAAAA)n"	
		join(<6480..6749,7602..7726,7811..7848,7929..8039,8152..8268,8487..8786)	
misc_feature	repeat_region	/gene="At2g23820"	
		<6480..8786	
		/note="T29E15.2"	
		join(6480..6749,7602..7726,7811..7848,7929..8039,8152..8268,8487..8557)	
		/gene="At2g23820"	
misc_feature	repeat_region	/note="unknown protein"	
		/codon_start=1	
		/protein_id="AAC63656.1"	
		/db_xref="GI:3738315"	
		/translation="MNKNRVFCCKLSLVFLVPSFTRSHIRFTYSNAGASSPNRAIHKMASDPQSGDVSPSPNVAAPSSSSSSAIDFLSLCTRLKTPPRAGWTKRDVKDESTADHMLKALISDDIPGVNRDKCMKMAIVHDAEAIVGDITPSCGISKEENRRESEALHMKLLGGGRAKEIAELWREYBENSPEAKVVKDFDKVELLQALLEYEOKFOTNICKAWASEVSRKKQH"	
misc_feature	repeat_region	complement(6654..6712)	
		/rpt_family="(GAA)n"	
		join(<9531..9582,9709..9779,9866..9947,10054..10169,10268..10396)	
		/gene="At2g23830"	
		<9531..10396	
misc_feature	repeat_region	/gene="At2g23830"	
		/note="T29E15.3"	
		join(9531..9582,9709..9779,9866..9947,10054..10169,10268..10396)	

repeat_region	repeat_region	/gene="At2g23830"	
		/note="unknown protein"	
		/codon_start=1	
		/protein_id="AAC63657.1"	
		/db_xref="GI:3738316"	
repeat_region	repeat_region	/translation="MSNNELLEIEPMYLQFFELKQKSCSLYLTNKTNNVAEKVKTTRNMGVRPNYGLILPKSTCKVLVMOAQRPVSDMQSFKEFKIQSYLASPGVTAKEVTRMFSGHVVVEETKLRVYVCSTTTNTSSPRTRRGFIQCF"	
		complement(11332..11371)	
		/rpt_family="(TAAA)n"	
		15801..15917	
		/rpt_family="(TAAA)n"	
repeat_region	repeat_region	15853..15918	
		/rpt_family="(CAAAA)n"	
		16051..16136	
		/rpt_family="(TAAA)n"	
		complement(join(<20442..20537,20617..20691,20894..20948,21145..21485))	
repeat_region	repeat_region	/gene="At2g23840"	
		complement(<20442..>21485)	
		/gene="At2g23840"	
		/note="T29E15.4"	
		complement(join(20442..20537,20617..20691,20894..20948,21145..21485))	
repeat_region	repeat_region	/gene="At2g23840"	
		/note="unknown protein"	
		/codon_start=1	
		/protein_id="AAC63658.1"	
		/db_xref="GI:3738317"	
repeat_region	repeat_region	/translation="MAHFSARKLLSSNDGLSFGVDSRQFQSLYMAGGSPVPIRMSRLKFCARIFSSYFPYKRKKIRDLONTENKLDIDEDNEDWDFDGGDDGLETDHLSFCRLGLDISVPVNVCMKRAICLEYMDKADVLEYVDQTVSSPTGSGFYIPAVLRVPHLLQVKKRRVKSLSRKNILLRDDYTCQ"	
		complement(22086..22781)	
		/gene="At2g23850"	
		/note="T29E15.5"	
		/pseudo	
repeat_region	repeat_region	join(<22999..23070,23189..23296,23402..>23554)	
		/gene="At2g23860"	
		<22999..>23554	
		/gene="At2g23860"	
		/note="T29E15.6"	
repeat_region	repeat_region	join(22999..23070,23189..23296,23402..23554)	
		/gene="At2g23860"	
		/note="unknown protein"	
		/codon_start=1	
		/protein_id="AAC63659.1"	
repeat_region	repeat_region	/db_xref="GI:3738318"	
		/translation="MTNPNKYVRPNYGLILPKSTCKLKRKPGSPDMQSNKEKFMIORVKASPAVTAKEVLTETFNKESGHLVEETKLRVYVCSTTTNITSPPTRNGLYFNVFVRHFLFTSL"	
		27493..27571	
		/rpt_family="(TAAA)n"	
		complement(29076..29146)	
repeat_region	repeat_region	/rpt_family="POLY_A"	
		29302..29885	
		/gene="At2g23870"	
		/note="T29E15.7"	
		/pseudo	
repeat_region	repeat_region	31472..31519	
		/rpt_family="(TA)n"	
		complement(join(<31877..32176,32267..32407,32498..32893,33162..35713,35886..>36147))	
		/gene="At2g23880"	
		complement(<31877..>36147)	
repeat_region	repeat_region	/gene="At2g23880"	
		/note="T29E15.8"	
		/pseudo	
		31472..31519	
		/rpt_family="(TA)n"	
repeat_region	repeat_region	complement(join(<31877..32176,32267..32407,32498..32893,33162..35713,35886..>36147))	
		/gene="At2g23880"	
		complement(<31877..>36147)	
		/gene="At2g23880"	
		/note="T29E15.8"	

Query Match 3.2%; Score 44.6; DB 8; Length 110211;
 Best Local Similarity 63.6%; Pred. No. 1.5;
 Matches 68; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
 Qy 1282 gtggacaagccaagtatcttataataattattatagcattagattatctgttatgttg 1341

RESULT 15

		Query Match	3.2%	Score 43.8;	DB 8;	Length 366;			
		Best Local Similarity	66.3%;	Pred. No.	1.1;				
		Matches	63;	Conservative	0;	Mismatches	32; Indels	0; Gaps	0;
QY	1294	agtatcttataaatattatagcattgagtgatatctgtttatagtgtgtttttttattaagc	1353						
Dd	241	ATTATATTGTTAAATTTTTTATTATTATATATATATATATAATTAATTTATT	182						
QY	1354	tgtggaaataaaaaataattattaaiaaaaaaaaaaaa	1388						
Dd	181	TATATAAATATTAATAATTTTATTATAAAAAATTATA	147						

Search completed: February 28, 2002, 00:56:47
Job time: 5649 sec

THIS PAGE BLANK (US--)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2002, 23:59:59 ; Search time 113.25 seconds
(without alignments)
10507.439 Million cell updates/sec

Title: US-08-816-011f-36

Perfect score: 1388

Sequence: 1 atgtgtaataatcccgatc.....ttattaaaaaaaaaaaaa 1388

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_1101.*
1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1386.4	99.9	1388	17 AAT18168	CORK potassium cha
2	85	6.1	936	22 AAF58252	Oligonucleotide D1
3	85	6.1	936	22 AAF58254	Oligonucleotide D1
4	85	6.1	936	22 AAF58257	Oligonucleotide D1
5	85	6.1	936	22 AAF58259	Oligonucleotide D1
6	85	6.1	936	22 AAF58262	Oligonucleotide D2
7	85	6.1	938	22 AAF58255	Oligonucleotide D1
8	83	6.0	936	22 AAF58252	Oligonucleotide D1
9	83	6.0	936	22 AAF58254	Oligonucleotide D1
10	83	6.0	936	22 AAF58257	Oligonucleotide D1
11	83	6.0	936	22 AAF58259	Oligonucleotide D1

C 12	83	6.0	936	22 AAF58262	Oligonucleotide D2
C 13	83	6.0	938	22 AAF58255	Oligonucleotide D1
C 14	43.2	3.1	244	22 AAF58238	Oligonucleotide D1
C 15	42.4	3.1	244	22 AAF58238	Oligonucleotide D1
C 16	39.4	2.8	362	22 AAF5894	Novel human polyu
C 17	39.2	2.8	440	18 AAT67765	H. pylori secreted
C 18	39.2	2.8	440	18 AAT77445	H. pylori secreted
C 19	38.6	2.8	310	22 AAH71505	Human cervical can
C 20	38.6	2.8	1558	17 AAT28255	Survival motor neu
C 21	38.6	2.8	1560	17 AAT18828	Human survival mot
C 22	38.6	2.8	1582	17 AAT28259	Survival motor neu
C 23	38.6	2.8	1582	17 AAT18831	Human survival mot
C 24	38.4	2.8	2439	20 AAH80667	Clone am856_3 enco
C 25	38.4	2.8	10732	21 AAA10594	Gene encoding a su
C 26	38	2.7	3499	21 AA255832	Staphylococcus aur
C 27	38	2.7	4709	18 AAV74898	Staphylococcus aur
C 28	37	2.7	2079	21 AA26360	Human secreted pro
C 29	36.8	2.7	65792	22 AAF28544	Genomic fragment #
C 30	36.6	2.6	513445	22 AA161373	Soybean 318013 reg
C 31	36.4	2.6	1338	21 AAC93411	Human secreted pro
C 32	36.2	2.6	1667	22 AAF85681	Pea blight resista
C 33	36.2	2.6	1851	21 AAC79950	Human secreted pro
C 34	36.2	2.6	3076	22 AAH46951	Human secreted pro
C 35	36.2	2.6	5761	22 AA159388	Human polynucleoti
C 36	36	2.6	1077	19 AAV39085	Human tumour necro
C 37	36	2.6	1077	21 AAA37772	Human polynucleoti
C 38	36	2.6	1164	19 AAV07654	Nucleotide sequenc
C 39	36	2.6	1179	22 AAH33171	Human colon cancer
C 40	36	2.6	2215	15 AAQ70882	Tomato spotted wil
C 41	36	2.6	2621	16 AAQ76109	Tomato spotted wil
C 42	36	2.6	2916	12 AAQ11655	TSWV S RNA. Tomat
C 43	36	2.6	2916	22 AAC89646	Tomato spotted wil
C 44	36	2.6	2964	22 AAF88022	Cucumber LBLOX DNA
C 45	36	2.6	2970	20 AAX86461	RNA encoding non-s

ALIGNMENTS

RESULT 1

AAT18168
ID AAT18168 standard; cDNA: 1388 BP.

XX
AC AAT18168;

XX
DT 16-OCT-1996 (first entry)

XX
DE CORK potassium channel gene.

XX
KW CORK; potassium channel; nematode; polyadenylation site;

XX
KW potassium dependence; Saccharomyces cerevisiae; potassium-agonist;

XX
KW potassium-antagonist; drug screening; nematocide; anthelmintic;

XX
KW cardiac disorder; ss.

XX
OS Caenorhabditis elegans.

XX
FT Key Location/Qualifiers

XX
FT CDS 1..1305

XX
FT /tag= a

XX
FT /product= CORK potassium channel

XX
FT polyA_signal 1359..1364

XX
FT /*tag= b

XX
PN WO9613520-A1.

XX
PD 09-MAY-1996.

XX
PF 25-OCT-1995; 95WO-US14364.

XX
PR 31-OCT-1994; 94US-0332312.

XX
PA (AMCY) AMERICAN CYANAMID CO.

	Query Match	6.1%	Score 85;	DB 22;	Length 936;
	Best Local Similarity	0.9%;	Pred. No. 8e-13;		
	Matches	7;	Conservative 431;	Mismatches 301;	Indels 0; Gaps 0;
Qy	650	ttttctttatcacaaagcaagattttaccactatcacacatcaaaaaagaatggaaattc	709		
Db	5	ttttctttatcacaaagcaagattttaccactatcacacatcaaaaaagaatggaaattc	64		
Qy	710	gcgaaagcggaacacgacagacgctctccatccattttggaccacattcacaaact	769		
Db	65	gcgaaagcggaacacgacagacgctctccatccattttggaccacattcacaaact	124		
Qy	770	gltatggcaactcttcaatgttggttctgttgcgttactctcaaatcttccctg	829		
Db	125	gltatggcaactcttcaatgttggttctgttgcgttactctcaaatcttccctg	184		
Qy	830	ttatgatgcggttaccactcgtgagagattcggcttccataacaaaattatgctcga	889		
Db	185	ttatgatgcggttaccactcgtgagagattcggcttccataacaaaattatgctcga	244		
Qy	890	acgatgaattctacactttgctcacaagttctctcgttcttcaatttgcgtgcgattg	949		
Db	245	acgatgaattctacactttgctcacaagttctctcgttcttcaatttgcgtgcgattg	304		
Qy	950	gatccatagttgcttccaagattcactgcgcgacaccccggtaccctcaaatgtgccata	1009		
Db	305	gatccatagttgcttccaagattcactgcgcgacaccccggtaccctcaaatgtgccata	364		
Qy	1010	tcttgcgtgctcttttcaattccattcttcttctcgaactatcgttccagacgcgtg	1069		
Db	365	tcttgcgtgctcttttcaattccattcttcttctcgaactatcgttccagacgcgtg	424		
Qy	1070	cttatcctgttttcttcttgagctactgacatttttgatggttggaattgccatgtctt	1129		
Db	425	cttatcctgttttcttcttgagctactgacatttttgatggttggaattgccatgtctt	484		
Qy	1130	tttcacatggatacctcagcgcctcgtgcaatgggatacacactccaaacgctcgtgccatctc	1189		
Db	485	tttcacatggatacctcagcgcctcgtgcaatgggatacacactccaaacgctcgtgccatctc	544		
Qy	1190	actactcaagatttgcgcgtcagcttctcgttttgcactcttatggttgcccttccacgc	1249		
Db	545	actactcaagatttgcgcgtcagcttctcgttttgcactcttatggttgcccttccacgc	604		
Qy	1250	gtggcctgtggcccggttgttattgagcactcgttggacaagccaagtatcttataataat	1309		
Db	605	gtggcctgtggcccggttgttattgagcactcgttggacaagccaagtatcttataataat	664		
Qy	1310	ttatgacttagatgatactgttattgtttttatttaagctgtgggaataaaataat	1369		
Db	665	ttatgacttagatgatactgttattgtttttatttaagctgtgggaataaaataat	724		
Qy	1370	tattaaaaaataaaaaa	1388		
Db	725	tattaaaaaataaaaaa	743		

7 RESULT

RESULT
AAF58255

AAF58255 standard; DNA: 938 BP.

XX
DT
CCZOC JMW

AC AAF58255;

XX
XX
155700 1111

DT 24-APR-2001 (first entry)

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

DE Oligonucleotide D1876.

XX

KW Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

XX

OS Synthetic.

١٢

XX	WO200107665-A2.
PN	
XX	
XX	01-FEB-2001.
PD	
XX	
XX	
XX	26-JUL-2000; 2000WO-US20476.
PF	
XX	
XX	26-JUL-1999; 99US-0145695.
PR	
PR	17-MAR-2000; 2000US-0190259.
XX	
XX	
XX	(CLIN-) CLINICAL MICRO SENSORS INC.
PA	
XX	
XX	Umek RM;
XX	
XX	WPI; 2001-159728/16.
DR	
XX	
XX	
XX	Nucleic acids containing electron-transfer group, useful as labels in
PT	hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT	a single surface
PT	
XX	
XX	Example 6; Page 127; 159pp; English.
PS	
XX	
CC	The present invention relates to a composition comprising two nucleic
CC	acids each containing an electron-transfer group (ETM) having
CC	different redox potentials. The invention is used for electronic
CC	detection of nucleic acids, especially of substitutions (mismatches)
CC	and single-nucleotide polymorphisms, e.g. for genotyping,
CC	monitoring gene expression.
XX	
XX	Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;
SQ	

Query Match 6.1%; Score 85; DB 22; Length 938;
Best Local Similarity 0.9%; Pred. No. 8e-13;
Matches 7: Conservative 431; Mismatches 301; Indels 0; Gaps 0;

650 ttttcttatcaacaagcagtattctaccactatcacccatcaaaaagggaattggaatttc 709
Dbb :
5 www
710 gcgaaaggcggaacgcagacggctctccacctcttgtgacacattcacaaact 769
Dbb :
65 www
770 gttatgggaactcttcaaattgttgcttgcgttactctcacaaattctccctg 829
Dbb :
125 www
830 ttatgatgcggttaccaactctggagattccggtctcatacaaaaattatctgtgaa 899
Dbb :
185 www
890 acgatgaaattacactttgctcacaaatttctcgtcttccaatttgttgcgtgcgattg 949
Dbb :
245 www
950 gatcatagttgcttccaagattcaactggcgacacccggttaacctcaaatgtgccata 1009
Dbb :
305 www
1010 tcttgctgctcttttcttaccattcttcttcttgcgaactatggtgccagcogtg 1069
Dbb :
365 www
1070 ctatatcctgttttttgagtctaactgacaaattttgtgattggtggaattgcgaatgttt 1129
Dbb :
425 www
1130 ttccaatgatacctcagcgctctggcaatgggatacactccaaacgctcgtgccatctc 1189
Dbb :
485 www
1190 actactcaaatattgcccgttcagcttccgttttgcaactcttatgttgcttccacgg 1249


```
QY 942 tgcgattgacatagttgcttccaagattcactggcgacacccgcttacctcaaat 1001
DB 477 WWWWWW: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1002 tgccataactctgctgctcttttcaattccattctcttctcgaactatcggtcca 1061
DB 417 WWWWWW: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1062 gacgctgcttacctgtttcttgagctcactgacatttttgattggtggaattgc 1121
DB 357 WWWWWW: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1122 catgtcttttcacatggatactcagcgtctctgcaatgggatacacactccaaagtcgt 1181
DB 297 WWWWWW: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1182 gcatctcactactcaagatttgcgctcagcttctcgttttgcaactcttatggtggcct 1241
DB 237 WWWWWW: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1242 tctcaccgctgacctgtgcttctgattgagcacttctggaagccaagtattct 1301
DB 177 WWWWWW: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1302 ataaatattatagcattagatagatactgtttatgtttttatttaagctgtggaat 1361
DB 117 WWWWWW: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1362 aaaaataattataaaaaaa 1388
DB 57 WWWWWW: : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 11

AAF58259/c
ID AAF58259 standard; DNA; 936 BP.

XX AAF58259;

XX 24-APR-2001 (first entry)

XX Oligonucleotide D2004.

XX Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

XX Synthetic.

XX WO200107665-A2.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-US20476.

XX 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX Umek RM;

XX WPI; 2001-159728/16.

DR Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface

PS Example 6; Page 128; 159pp; English.

XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
XX

CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 6.0%; Score 83; DB 22; Length 936;

Best Local Similarity 0.7%; Pred. No. 2.7e-12;

Matches 5; Conservative 436; Mismatches 306; Indels 0; Gaps 0;

QY 642 aattgcactttcttattcacaaagcagattttaccactatcaccatcaaaaaggaat 701

DB 777 WWWWWW: :

QY 702 ggaattcgcgaaagcggaacacgacacgctcctccattctttggaccacatt 761

DB 717 WWWWWW: :

QY 762 cacaactgttatgggcaactcttcaattttgttctgttgcgttactctcacaat 821

DB 657 WWWWWW: :

QY 822 ctctcctgttatgacacgcttaccactcgtggagattccggtctcctcaaaaattat 881

DB 597 WWWWWW: :

QY 882 gtctgaaacgataaactcacactttgtctcacaagtttctcgtcttcaattgttcgc 941

DB 537 WWWWWW: :

QY 942 tgcgattggacatagttgcttccaagattcactggcgacaccccgcttacctcaaat 1001

DB 477 WWWWWW: :

QY 1002 tgccataactctgctgctcttttccattctcttctcgtcgaactatcggtcca 1061

DB 417 WWWWWW: :

QY 1062 gacgctgcttacctgtttcttgagctcactgacatttttgattggtggaattgc 1121

DB 357 WWWWWW: :

QY 1122 catgtcttttcacatggatactcagcgtctctgcaatgggatacacactccaaagtcgt 1181

DB 297 WWWWWW: :

QY 1182 gcatctcactactcaagatttgcgctcagcttctcgttttgcaactcttatggtggcct 1241

DB 237 WWWWWW: :

QY 1242 tctcaccgctgacctgtgcttctgattgagcacttctggaagccaagtattct 1301

DB 177 WWWWWW: :

QY 1302 ataaatattatagcattagatagatactgtttatgtttttatttaagctgtggaat 1361

DB 117 WWWWWW: :

QY 1362 aaaaataattataaaaaaa 1388

DB 57 WWWWWW: :

RESULT 12

AAF58262/c

ID AAF58262 standard; DNA; 936 BP.

XX AAF58262;

XX 24-APR-2001 (first entry)

XX Oligonucleotide D2007.

XX

KW Electron-transfer group; ETM; mismatch; genotyping;
 KW gene expression; ss.
 XX Synthetic.
 OS WO200107665-A2.
 PN 01-FEB-2001.
 XX 26-JUL-2000; 2000WO-US20476.
 XX 26-JUL-1999; 99US-0145695.
 PR 17-MAR-2000; 2000US-0190259.
 XX (CLIN-) CLINICAL MICRO SENSORS INC.
 PA Umek RM;
 XX WPI; 2001-159728/16.
 DR Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface
 XX Example 6; Page 128; 159pp; English.
 PS The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.
 XX Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;
 SQ

Query Match 6.0%; Score 83; DB 22; Length 936;
 Best Local Similarity 0.7%; Pred. No. 2.7e-12;
 Matches 5; Conservative 436; Mismatches 306; Indels 0; Gaps 0;

Qy 642 aatgcactttcttaccagagatttctaccactaccatcaaaaggaat 701
 Db 777 WWWWWW
 Qy 702 ggaattcggaagcgagaccgctctccattcttggaccacatt 761
 Db 717 WWWWWW
 Qy 762 cacaaactgttggaactcttcaatgttctgtctgtcgttccgttactctcaaat 821
 Db 657 WWWWWW
 Qy 822 ctccctgttatgatgaccgtttaccactcgtggagattccggtctctaaacaaattat 881
 Db 597 WWWWWW
 Qy 882 gtctgaaacagatgaaattctacatttctgtcacaagtttctcgtcttcaattgttcgc 941
 Db 537 WWWWWW
 Qy 942 tgcgattgacccatagttgttccagattcactgcccagaccccggttaccctcaaat 1001
 Db 477 WWWWWW
 Qy 1002 tgcataatctgcgtcttcttccattcttcttctctgtcactatcgttcca 1061
 Db 417 WWWWWW
 Qy 1062 gacgcgtgcttctcttcttctgtgactactgacatttttggattggtggaattgc 1121
 Db 357 WWWWWW
 Qy 1122 catgtcttttccatggataccctcagcgtctgtggcaatgggatacacctccaaacgtcgt 1181

Db 297 WWWWWW
 Qy 1182 gccatctcactactcaagatttgcgcctcagcttccgttttgcaactcttattggttgccct 1241
 Db 237 WWWWWW
 Qy 1242 tctcaccggtgacctgtggccgttcttattgagcactctggaacagcaagtattctt 1301
 Db 177 WWWWWW
 Qy 1302 ataaatttatagcattagatgactattgtttattgttttatttaagctgtggaat 1361
 Db 117 WWWWWW
 Qy 1362 aaataattattaaaaa 1388
 Db 57 WWWWWW

RESULT 13
 AAF58255/C
 ID AAF58255 standard; DNA; 938 BP.
 XX AAF58255;
 XX 24-APR-2001 (first entry)
 XX Oligonucleotide D1876.
 XX Electron-transfer group; ETM; mismatch; genotyping;
 KW gene expression; ss.
 XX Synthetic.
 XX WO200107665-A2.
 XX 01-FEB-2001.
 XX 26-JUL-2000; 2000WO-US20476.
 XX 26-JUL-1999; 99US-0145695.
 PR 17-MAR-2000; 2000US-0190259.
 XX (CLIN-) CLINICAL MICRO SENSORS INC.
 XX Umek RM;
 XX WPI; 2001-159728/16.
 XX Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface
 XX Example 6; Page 127; 159pp; English.
 XX The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.
 XX Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 6.0%; Score 83; DB 22; Length 938;
 Best Local Similarity 0.7%; Pred. No. 2.7e-12;
 Matches 5; Conservative 436; Mismatches 306; Indels 0; Gaps 0;

Qy 642 aatgcactttcttaccagagatttctaccactaccatcaaaaggaat 701
 Db 777 WWWWWW

[illegible]

RESULT	14	
AAF58238/C		
ID	AAF58238 standard; DNA; 244 Bp.	
XX		
XX		
AC	AAF58238;	
XX		
DT	24-APR-2001 (first entry)	
XX		
DE	Oligonucleotide D1250:D1102.	
XX		
KW	Electron-transfer group; ETM; mismatch; genotyping;	
KW	gene expression; ss.	
XX		
OS	Synthetic.	
XX		
PN	WO200107665-A2.	
XX		
PD	01-FEB-2001.	
XX		
PF	26-JUL-2000; 2000WO-US20476.	
XX		
PR	26-JUL-1999; 99US-0145695.	
PR	17-MAR-2000; 2000US-0190259.	

```

XX      (CLIN-) CLINICAL MICRO SENSORS INC.
PA      Umek RM;
PI      WPI; 2001-159728/16.
PP      Nucleic acids containing electron-transfer group, useful as labels in
PR      hybridization assays, e.g. for genotyping, allowing repeat analyses on
PS      a single surface
XX      Example 4; Page 120; 159pp; English.
XX      The present invention relates to a composition comprising two nucleic
CC      acids each containing an electron-transfer group (ETM) having
CC      different redox potentials. The invention is used for electronic
CC      detection of nucleic acids, especially of substitutions (mismatches)
CC      and single-nucleotide polymorphisms, e.g. for genotyping,
CC      monitoring gene expression.
XX      Sequence 244 BP; 19 A; 9 C; 12 G; 10 T; 194 other:
SQ
Query Match          3.1%;   Score 43.2; DB 22; Length 244;
Best Local Similarity 8.0%; Pred. No. 0.057;
Matches 12; Conservative 95; Mismatches 43; Indels 0; Gaps 0;

Qy 1239 ccttctaaccggtggcctgtgcccgttgtattaggacacttcgtggacaagccaagtat 1298
    ||| | |||| | | : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 226 CCTCATTGATGGTGTCTTTTAACAAWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 167

Qy 1299 cttaataatttatagcatgagtactgtttatatgtttttttaagctgtgg 1358
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 166 WWWWWWWWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 107

Qy 1359 aataaaataattataaaaaaaaataaaaaa 1388
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 106 WGWMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM 77

```

RESULT	15
AAF58238	
ID	AAF58238 standard; DNA; 244 BP.
XX	
AC	AAF58238;
XX	
DT	24-APR-2001 (first entry)
XX	
DE	Oligonucleotide D1250:D1102.
XX	
KW	Electron-transfer group; ETM; mismatch; genotyping;
KW	gene expression; ss.
XX	
OS	Synthetic.
XX	
PN	WO200107665-A2.
XX	
PD	01-FEB-2001.
XX	
PF	26-JUL-2000; 2000WO-US20475.
XX	
PR	26-JUL-1999; 99US-0145695.
PR	17-MAR-2000; 2000US-0190259.
XX	
PA	(CLIN-) CLINICAL MICRO SENSORS INC.
XX	
PI	Umek RM;
XX	
DR	WPI; 2001-159728/16.
XX	
PT	Nucleic acids containing electron-transfer group, u
PT	hybridization assays, e.g. for genotyping, allowing
PT	a single surface


```

XX Example 4; Page 120; 159pp; English.
PS
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETW) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SO Sequence 244 BP; 19 A; 9 C; 12 G; 10 T; 194 other;

Query Match
Best Local Similarity 3.1%; Score 42.4; DB 22; Length 244;
Matches 1; Conservative 96; Mismatches 27; Indels 0; Gaps 0;

Qy 1265 ttgtattgagcacttcgtggacacgaagtcatttataatatttagcattagagt 1324
Db 17 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 76

Qy 1325 atacttattatattgttttttatttaagctgtggaataataattattataaaaaaaa 1384
Db 77 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 136

Qy 1385 aaaa 1388
Db 137 www 140

```

Search completed: February 28, 2002, 00:52:31
Job time: 3152 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2002, 23:21:59 ; Search time 1212.08 seconds
(without alignments)
12305.409 Million cell updates/sec

Title: US-08-816-011f-36
Perfect score: 1388
Sequence: 1 atggtaataatcagcagatc.....ttattataaaaaaaaaa 1388

Scoring table: IDENTITY_NUC
Gap 10.0, Gapext 1.0
Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Existing first 45 summaries

Database : EST:
1: em_estfun:
2: em_esthum:
3: em_estin:
4: em_estom:
5: em_estpl:
6: em_estba:
7: em_estro:
8: em_estov:
9: em_htc:
10: gb_estl:
11: gb_est2:
12: gb_htc:
13: gb_gss:
14: em_gss_fun:
15: em_gss_hum:
16: em_gss_inv:
17: em_gss_pln:
18: em_gss_pro:
19: em_gss_rod:
20: em_gss_vrt:
21: em_gss_other:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	86.6	6.2	533	10 AU222343	AU222343 AU222343
c 2	77.6	5.6	332	10 AU201199	AU201199 AU201199
c 3	75.6	5.4	710	10 AU216992	AU216992 AU216992
4	75.2	5.4	648	10 AU453425	AW453425 SWOV3MCA
5	73.8	5.3	327	10 AU109268	AU109268 AU109268
6	73.8	5.3	642	10 AU1082929	AI082929 SWAMC12
7	70.8	5.1	495	10 AU209436	AU209436 AU209436
8	70.2	5.1	360	11 C13829	C13829 C13829 Yuji
c 9	67.8	4.9	300	10 AU113471	AU113471 AU113471
c 10	57.6	4.1	300	10 AU115402	AU115402 AU115402
c 11	57.4	4.1	300	10 AU114995	AU114995 AU114995
12	56.2	4.0	997	13 CNS0057E	AL060767 Drosophil

13	54.4	3.9	258	10 AA406898	AA406898
14	53	3.8	656	10 AW409482	AW409482
15	50.4	3.6	345	10 AU109592	AU109592
16	49	3.5	606	10 AA948920	AA948920
c 17	48.8	3.5	901	13 CNS0760F	CNS0760F
18	48.2	3.5	1201	13 CNS0010J	CNS0010J
19	47.6	3.4	274	10 AA842090	AA842090
20	47	3.4	228	10 AI784877	AI784877
21	46.8	3.4	729	10 AI239052	AI239052
22	46.2	3.3	360	11 R03450	R03450
23	46.2	3.3	373	11 C43216	C43216
24	45.8	3.3	274	10 BB181657	BB181657
25	45.6	3.3	350	10 AU111089	AU111089
26	45.6	3.3	351	10 AU111161	AU111161
27	45.6	3.3	738	13 AQ842781	AQ842781
c 28	45.6	3.3	759	13 AQ254298	AQ254298
29	45	3.2	375	11 C44886	C44886
30	44.6	3.2	340	10 AU111560	AU111560
31	44.6	3.2	340	10 AU111734	AU111734
c 32	44.6	3.2	510	13 AQ254261	AQ254261
c 33	44.4	3.2	429	10 AW829621	AW829621
c 34	44.4	3.2	448	10 AW829541	AW829541
c 35	44.2	3.2	333	10 AI431074	AI431074
c 36	44.2	3.2	441	11 BF459052	BF459052
37	44.2	3.2	459	10 AI595433	AI595433
38	44.2	3.2	475	10 AA058117	AA058117
c 39	44.2	3.2	566	10 AW333515	AW333515
c 40	44.2	3.2	626	10 AW048063	AW048063
41	44.2	3.2	1101	13 CNS012EM	CNS012EM
42	44.2	3.2	2664	12 AK005136	AK005136
c 43	43.6	3.1	227	11 BF460904	BF460904
44	43.6	3.1	234	10 BB169820	BB169820
45	43.6	3.1	327	10 BB182391	BB182391

ALIGNMENTS

RESULT 1
AU222343/c
LOCUS AU222343 533 bp mRNA EST 17-JUL-2001
DEFINITION AU222343 unpublished oligo-capped cDNA library, stage L1
Caenorhabditis elegans cDNA clone yk1015b04 3', mRNA sequence.
ACCESSION AU222343
VERSION AU222343.1 GI:14860500
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 533)
AUTHORS Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2001)
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
FEATURES
source
1..533
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk1015b04"
/clone_lib="unpublished oligo-capped cDNA library, stage L1"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"

and DNA pol I. The library was constructed in the lambda Uni-zap XR vector and has 1 x 10⁶ independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigman@ybc.org).

BASE COUNT	168	a	126	c	120	g	234	t
ORIGIN								

	Query Match	5.4%	Score 75.2	DB 10	Length 648
	Best Local Similarity	48.9%	Pred. No. 1.5e-05		
	Matches 305	Conservative 0	Mismatches 298	Indels 21	Gaps 3
Qy	30	tgcgcgttgagcaggaagcattcccaagagacaaagtacaataattgtctactggtcgctcat	89		
Db	28	TGATATGGTGTTCGACAAACCTCCCAAGACAAGTACAATGCTCTTTATTTCATTTTACT	87		
Qy	90	tcttggttggaattcggaggttctctgcgcattggaataatttctactactcgcctcgagta	149		
Db	88	ACTTCACGGGATAGCGCGTCTTGATGCCGTGGAACATGTTATCATCAAAATTGCTCCATCTTA	147		
Qy	150	ttatgtgaattcatctgggttc-----aaaccggatggcggtggagacatgg-----	194		

	Query Match	5.4%	Score 75.2	DB 10	Length 648
	Best Local Similarity	48.9%	Pred. No. 1.5e-05		
	Matches 305	Conservative 0	Mismatches 298	Indels 21	Gaps 3
Qy	30	tgcgcgttgagcaggaagcattcccaagagacaaagtacaataattgtctactggtcgctcat	89		
Db	28	TGATATGGTGTTCGACAAACCTCCCAAGACAAGTACAATGCTCTTTATTTCATTTTACT	87		
Qy	90	tcttggttggaattcggaggttctctgcgcattggaataatttctactactcgcctcgagta	149		
Db	88	ACTTCACGGGATAGCGCGTCTTGATGCCGTGGAACATGTTATCATCAAAATTGCTCCATCTTA	147		
Qy	150	ttatgtgaattcatctgggttc-----aaaccggatggcggtggagacatgg-----	194		

	Query Match	5.4%	Score 75.2	DB 10	Length 648
	Best Local Similarity	48.9%	Pred. No. 1.5e-05		
	Matches 305	Conservative 0	Mismatches 298	Indels 21	Gaps 3
Qy	30	tgcgcgttgagcaggaagcattcccaagagacaaagtacaataattgtctactggtcgctcat	89		
Db	28	TGATATGGTGTTCGACAAACCTCCCAAGACAAGTACAATGCTCTTTATTTCATTTTACT	87		
Qy	90	tcttggttggaattcggaggttctctgcgcattggaataatttctactactcgcctcgagta	149		
Db	88	ACTTCACGGGATAGCGCGTCTTGATGCCGTGGAACATGTTATCATCAAAATTGCTCCATCTTA	147		
Qy	150	ttatgtgaattcatctgggttc-----aaaccggatggcggtggagacatgg-----	194		

	Query Match	5.4%	Score 75.2	DB 10	Length 648
	Best Local Similarity	48.9%	Pred. No. 1.5e-05		
	Matches 305	Conservative 0	Mismatches 298	Indels 21	Gaps 3
Qy	30	tgcgcgttgagcaggaagcattcccaagagacaaagtacaataattgtctactggtcgctcat	89		
Db	28	TGATATGGTGTTCGACAAACCTCCCAAGACAAGTACAATGCTCTTTATTTCATTTTACT	87		
Qy	90	tcttggttggaattcggaggttctctgcgcattggaataatttctactactcgcctcgagta	149		
Db	88	ACTTCACGGGATAGCGCGTCTTGATGCCGTGGAACATGTTATCATCAAAATTGCTCCATCTTA	147		
Qy	150	ttatgtgaattcatctgggttc-----aaaccggatggcggtggagacatgg-----	194		

	Query Match	5.4%	Score 75.2	DB 10	Length 648
	Best Local Similarity	48.9%	Pred. No. 1.5e-05		
	Matches 305	Conservative 0	Mismatches 298	Indels 21	Gaps 3
Qy	30	tgcgcgttgagcaggaagcattcccaagagacaaagtacaataattgtctactggtcgctcat	89		
Db	28	TGATATGGTGTTCGACAAACCTCCCAAGACAAGTACAATGCTCTTTATTTCATTTTACT	87		
Qy	90	tcttggttggaattcggaggttctctgcgcattggaataatttctactactcgcctcgagta	149		
Db	88	ACTTCACGGGATAGCGCGTCTTGATGCCGTGGAACATGTTATCATCAAAATTGCTCCATCTTA	147		
Qy	150	ttatgtgaattcatctgggttc-----aaaccggatggcggtggagacatgg-----	194		

	Query Match	5.4%	Score 75.2	DB 10	Length 648
	Best Local Similarity	48.9%	Pred. No. 1.5e-05		
	Matches 305	Conservative 0	Mismatches 298	Indels 21	Gaps 3
Qy	30	tgcgcgttgagcaggaagcattcccaagagacaaagtacaataattgtctactggtcgctcat	89		
Db	28	TGATATGGTGTTCGACAAACCTCCCAAGACAAGTACAATGCTCTTTATTTCATTTTACT	87		
Qy	90	tcttggttggaattcggaggttctctgcgcattggaataatttctactactcgcctcgagta	149		
Db	88	ACTTCACGGGATAGCGCGTCTTGATGCCGTGGAACATGTTATCATCAAAATTGCTCCATCTTA	147		
Qy	150	ttatgtgaattcatctgggttc-----aaaccggatggcggtggagacatgg-----	194		

JOURNAL COMMENT

Unpublished (2001)
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES

Source
1. 495
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk1015b04"
/clone_lib="unpublished oligo-capped cDNA library, stage 1"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
BASE COUNT 146 a 80 c 98 g 171 t
ORIGIN

Query Match 5.1%; Score 70.8; DB 10; Length 495;
Best Local Similarity 51.0%; Pred. NO. 0.00013;
Matches 202; Conservative 0; Mismatches 182; Indels 12; Gaps 1;

QY 39 gcagaagcatttcgaagagacaaatgataattgtctactggtcgctcattctgttgg 98
DB 110 GAAAGAGAGTTCCACCAACAGATCGATGTTATCTTTTAAATTTTACAAATGACAGG 169
QY 99 attcggagttcttgcgaatgaatgttcaattactatcgccctgagattatgtgaa 158
DB 170 AATGGGAATTTAATGTCATGAATATGTTTATACAAATGTCACAAATATACCA 229
QY 159 ttattggttcaaacggatggtgagacatggtattcgaagaattcattggtattt 218
DB 230 TTATTGGTTTAAACA-----ATACAAATTTATCAAGACAGTTTATGTCATTAAT 277
QY 219 gacgattggctcaacacttcaaacgaagcaatttaatttttcaaccgttctcattat 278
DB 278 TGGAGTGACGTACAGATTCACAAATGTTGGAATATGATTTTAAATACAAATTTGGTCA 337
QY 279 tctggtccctgatctacccggtcttctgctcgggttcttcaacatcgcaacctgac 338
DB 338 GGTAGTTCATGATGCTTCGAGTTGTTGTTCTCTCATTTGGAATGCAATTTTGATTGG 397
QY 339 aatcattctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 398
DB 398 GGTCAATGTTATCTTGGCAATTTTGTAAACACCATCTCCAGACAGTTTACATGTTTAA 457
QY 399 ctgggttaactcttgggaatggcgacttcaatcaattt 434
DB 458 TATTGTTACTCTAATCATCATTAATGGCTATGAATTT 493

RESULT 8
LOCUS C13829
DEFINITION C13829 Yuji Kohara unpublished cDNA Caenorhabditis elegans CDNA clone yk183c7 5', mRNA sequence.
ACCESSION C13829
VERSION C13829.1 GI:1561382
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.

REFERENCE
AUTHORS Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano, M., Miyata, A. and Nishigaki, A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)

COMMENT

Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES

Source
1. 360
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk183c7"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT 79 a 107 c 71 g 99 t 4 others
ORIGIN

Query Match 5.1%; Score 70.2; DB 11; Length 360;
Best Local Similarity 56.0%; Pred. NO. 0.00019;
Matches 149; Conservative 0; Mismatches 116; Indels 1; Gaps 1;

QY 911 tcacaagttctctcttctcaattgttcgtcgatggtgattggtccatagttgcttcca-ag 969
DB 29 TTANACATCTCTTCAATTCACAGTGTGTTGGCTTCATCGATCAATCGTCGCGGGAGNA 88
QY 970 attcactggcgacaccccggttacctcaaatgtgcataattgtcgatgctcttttcatt 1029
DB 89 AAGCAATGGCGCGCCGGAACAGCTCTGGATCCAGCTACCTCGTCTCCTCTACATC 148
QY 1030 caattctcttcttctgaaatctggttcagagcggtgtattcctcttcttcttgg 1089
DB 149 CCAATCTTTCATCTTGCACATACCTCCCTGAGACCCGACACTCCCGGTCTTCTCGAA 208
QY 1090 tctactgacatttctgtgattggtgaaattgcatgtcttcttccatggtacacctcagc 1149
DB 209 TCCAGTGGCTTTTCATCTATGNGCCGATCGATGAGCTTCGGAAGTGATCTCTCA 268
QY 1150 gctctggcaatgggatacactccaaa 1175
DB 269 GGACTCTCCATGATGATACACTTCAA 294

RESULT 9
LOCUS AU113471/c
DEFINITION AU113471 unpublished oligo-capped cDNA library Caenorhabditis elegans CDNA clone yk708a4 3', mRNA sequence.
ACCESSION AU113471
VERSION AU113471.1 GI:10927038
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.

REFERENCE
AUTHORS Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y. and Sugano, S.
TITLE A complementary view of the C. elegans genome
JOURNAL Unpublished (2000).
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES
Source
1. 300
/organism="Caenorhabditis elegans"

```

/strain="N2"
/db_xref="taxon:6239"
/clone="yk708a4"
/clone_lib="unpublished oligo-capped cDNA library"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT      114 a 50 c 59 g 77 t
ORIGIN

Query Match      4.9%; Score 67.8; DB 10; Length 300;
Best Local Similarity 55.8%; Pred. No. 0.00061;
Matches 129; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

Qy 1024 ttcttcattctctctctcgaactatcggtgcacagcgctgtcttcctctcttc 1083
Db 299 TTCATTCATTCTTCATGTTTCAGCAACTATTTACCACTTCCTCGTACTATGGGAGTATTA 240

Qy 1084 ttgagtcctactgacattttgtgattggtggaattgccatgtctttttcacatggatc 1143
Db 239 TTCACCAAGCAATGGATCTCTCTTTTGGTAACACATTTATTGGCATTACACAAGTGGATAT 180

Qy 1144 ctacgctctggaatgggatacactccaaacgctgctgcacatctcactcactcaagattt 1203
Db 179 TTCATCTCTTTGGGAATGATGATATACACACCGTGTCTGCCACCGGAATACTCAAAATTA 120

Qy 1204 gccctcagcttccgtttgcactctatggttggtcctcttcaccggtggc 1254
Db 119 GCTGCTCAAGTCTGCAATGCTCTCTGCTGTTGTTGTTACTGCTGTGTC 69

RESULT 10
AUI15402/c      300 bp mRNA EST 19-OCT-2000
LOCUS
DEFINITION
AUI15402 unpublished oligo-capped cDNA library Caenorhabditis
AUI15402
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans.
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
A complementary view of the C. elegans genome
Unpublished (2000)
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1. .300
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="yk732c11"
/clone_lib="unpublished oligo-capped cDNA library"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT      93 a 68 c 56 g 83 t
ORIGIN

Query Match      4.1%; Score 57.6; DB 10; Length 300;
Best Local Similarity 52.5%; Pred. No. 0.072;
Matches 126; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

```

```

Qy 1015 cgtgcttttccattccattcttcttctgcaactatcggtccagacgctgcttat 1074
Db 299 CGAACCGTATTCATTCCTATTTATCTTTCTGTAATATATCTGTCAGATAGAGATGG 240

Qy 1075 cctgttttcttgagtcactgacattttgtgattggtggaattgccatgtcttttca 1134
Db 239 CCTGTTGGTTTAAAGAAATGAATGGTGGTTCATATTGGATGTACCAATAATGGCATTTACT 180

Qy 1135 catgatacctcagcgtctggaatgggatacactccaaacgctgcgcatctcactac 1194
Db 179 TGTGGTTTATATGACGACTTTAGCGCTGATTTATACACCAAGCAAGTCCAGCTAGATAT 120

Qy 1195 tcaagatttgcgcctcagctctccgttttgcactcttattggttgcctctcaccggtggc 1254
Db 119 CAGAAGCTAAGTGAATGCTTGTCATCAATTTCTCTAATGCTCGGAATTCATCGGAGTC 60

RESULT 11
AUI14995/c      300 bp mRNA EST 19-OCT-2000
LOCUS
DEFINITION
AUI14995 unpublished oligo-capped cDNA library Caenorhabditis
AUI14995
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans.
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
A complementary view of the C. elegans genome
Unpublished (2000)
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1. .300
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="yk724f10"
/clone_lib="unpublished oligo-capped cDNA library"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT      96 a 67 c 53 g 84 t
ORIGIN

Query Match      4.1%; Score 57.4; DB 10; Length 300;
Best Local Similarity 49.5%; Pred. No. 0.079;
Matches 148; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

Qy 1022 ttcttcattccattcttcttctgcaactatcggtccagacgctgcttatcctgttt 1081
Db 300 TATTCATTCCTCATTTTCTCTGTAATATGCTTCAGATAGAGATGGCTGTTT 241

Qy 1082 tccttgagtcactgacattttgtgattggtggaattgccatgtcttttccatgat 1141
Db 240 GGTTTAAAGAAATGAATGGTGGTTCATTCATTCACCAATGGAATTTACTTGTGTTT 181

Qy 1142 acctcagcgtctggaatgggatacactccaaacgctgcgcatctcactcactcaagat 1201
Db 180 ATATGACGACTTTAGCGCTGATTTATACACCAAGCAAGTCCAGCTAGATATCAGAAGC 121

Qy 1202 ttgccgctcagcttccgtttgctcactcttattggttgccctctcaccggtgacctg 1261

```

```

Db 120 TAAGTGAATGCTGCATCAATTTCCCTAATGCTCGGAATTCCTATCGGAGTCGCCAGTA 61
Qy 1262 cggttgttatgagcaacttcggtggacaagccaagtatcttataaataattatagatta 1320
Db 60 CACCATTCGCTGCATGGCGGTGGAGTCTGATAGGAAGTAGAAGAACATGACTGTATAA 2

RESULT 12
CNS005TE 997 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL060767
KEYWORDS AL060767.1 GI:4943573
SOURCE GSS.
ORGANISM fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 997)
REFERENCE
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osogawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1..997
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR12K22"
/note="end : TET3"

BASE COUNT 89 a 99 c 13 g 258 t 538 others
ORIGIN

Query Match 4.0%; Score 56.2; DB 13; Length 997;
Best Local Similarity 18.6%; Pred. NO. 0.1;
Matches 116; Conservative 172; Mismatches 335; Indels 0; Gaps 0;

Qy 661 acaagcaagatttaccacattaccatcaaaaaggaaatcggaattcgcgaaaggcg 720
Db 356 ANNNAANNAANNNANNAANNAANNNNTNANNANNAANNAANNNNTATGNNNNN 415
Qy 721 gaaacgcagacgcgctccattcttggaccacattcacaaactgtatgggcaa 780
Db 416 NNNNNNNNNAANNAANNNANGCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 475
Qy 781 ctcttcaatgttggttcgttgcgttactctacaaatcttccctgttatgatgacc 840
Db 476 TTYTYYTYYTYYTCTYYCYCCYCYCCYCYCCYCYCCYCYCCYCYCCYCYCYCYCY 535
Qy 841 gttaccactcgtggagatcccggtctctctcaacaaaattatgtctgaaacgatgaa 900
Db 536 TYCTCYTYYTYYTYYTYYTYYTYYTYYTYYTYYTYYTYYTYYTYYTYYTYYTYY 595

```

```

Qy 901 tacatttgcgcacaagtttctcgtcttcaatttgcgtcgcgattgagatccatagtt 960
Db 596 CYTCTYTYCYCYCYCTYYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCY 555
Qy 961 gcttccaagattcactgcccgcacaccgcgttaccctcaaaatttgcctaaatcttgcgtgct 1020
Db 656 YVAYKYYCYCYCYCTYYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCY 715
Qy 1021 cttttcattccatcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1080
Db 716 YTCTCYYYYYTYYTYYTYYTYYTYYTYYTYYTYYTYYTYYTYYTYYTYYTYYTYY 775
Qy 1081 tctcttgagctactgacacatttctgtgattggtggaatgcccattctcttcttccatgga 1140
Db 776 YVCCYCYCYCTYYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCY 835
Qy 1141 tacctcagcgtctggaatgggatactacacacacacacacacacacacacacacacacac 1200
Db 836 YCTYCTCTYTYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCY 895
Qy 1201 ttgctcgtcagcttctcgttcttgcactcttcttgccttcttgccttcttgccttcttgcct 1260
Db 896 TYVCCYCYCCCTSYCYCYCTYYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCY 955
Qy 1261 ccgctgtgtattgagcacttcgt 1283
Db 956 TTYYYTYYTCTYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCY 978

RESULT 13
AA406898 258 bp mRNA EST 01-MAY-1997
LOCUS MBAPCZ7F07T3 Brugia malayi adult female cDNA (SAW96MLW-BmaF) Brugia
DEFINITION malayi cDNA clone AFCZ7F07 5', mRNA sequence.
ACCESSION AA406898
VERSION AA406898.1 GI:2064981
KEYWORDS EST.
SOURCE Brugia malayi.
ORGANISM Brugia malayi.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Brugia.
REFERENCE 1 (bases 1 to 258)
AUTHORS Blaxter,M.L., Waterfall,M., Daub,J., Lizotte,M., Baron,L. and Jones
,S.J.
TITLE Genes expressed in adult female Brugia malayi
JOURNAL Unpublished (1996)
COMMENT Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JT, UK
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The ABI trace of this sequence can be viewed at
http://www.sanger.ac.uk/brugia/AFC/MBAPCZ7F07T3.html
Seq primer: T3
Location/Qualifiers
1..258
/organism="Brugia malayi"
/db_xref="taxon:6279"
/clone="AFCZ7F07"
/clone_lib="Brugia malayi adult female cDNA (SAW96MLW-BmaF)"
/sex="female"
/dev_stage="adult"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Lymphatic filarial nematode parasite of humans.
mRNA was prepared from approximately 50 adult females.
Isolated from the peritoneal cavity of birds and
converted to double-stranded cDNA using reverse

```

transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 5 x 10⁶ independent recombinants and the average insert size is ~900bp. The library was constructed by Michelle Lizotte-Maniewski. The library is available from Dr. S.A. Williams, email: genome@smith.edu."

BASE COUNT	66 a	46 c	52 g	94 t
ORIGIN				

Query Match	3.98;	Score 54.4;	DB 10;	Length 258;
Best Local Similarity	61.18;	Pred. No. 0.33;		
Matches 88; Conservative	0;	Mismatches 56;	Indels 0	

Qy	42	qgaagcttccaaagacagaacagataatgctactagctcgctcatctctgttgatt	101
Db	19	GGATAAACCTCCCTAAAGACAACATATATGCTGTCTATTTTCATTTTACTACTCACGGAT	78
Qy	102	cgaagttctctgcactggaatgattcattactatgcgccctgagtatatgtgaatta	161
Db	79	AGGTGTTTGTATGCCATGGAAACATGTCATCAAAATGTCCCTCTTACTACGTGGATTA	138
Qy	162	ttggttcaaacccgatggcgtgga	185
Db	139	TAAATTTGTGGAGGTGAGCGCCA	162

RESULT	14
AW409482	
LOCUS	656 bp mRNA EST 08-FEB-2000
DEFINITION	SWOV3MCAM40A01SK Onchocerca volvulus molting L3 larva cDNA (SL56MLW-OvmL3) Onchocerca volvulus cDNA clone SWOV3MCAM40A01 5', mRNA sequence.
ACCESSION	AW409482
VERSION	AW409482.1 GI:6935020
KEYWORDS	EST.
SOURCE	Onchocerca volvulus. Onchocerca volvulus. Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
ORGANISM	Onchocercidae; Onchocerca. 1 (bases 1 to 656) Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S. Genes expressed in molting L3 larvae of Onchocerca volvulus Unpublished (1997)
REFERENCE	Contact: Steven A. Williams
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA

Tel: 4135853620
 Fax: 4135853786
 Email: genome@smith.edu
 Seq primer: pbluescript SK.
 Location/Qualifiers
 1 .. 656

FEATURES
SOURCE

```

/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOV3MCAM4 OA01"
/clone_lib="Onchocerca volvulus molting L3 larva cDNA
(SL96MLW-OvmL3)"
/dev_stage="molting L3"
/lab_host="XLI-Bio MRF"
/notes="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCIC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (mL3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H

```

and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10⁶ independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigman@bc.org)."

BASE COUNT	192 a	119 c	120 g	225 t
ORIGIN				

Query Match 3.8%; Score 53; DB 10; Length 656;
Best Local Similarity 46.7%;
Pred. No. 0.5;
Matches 245; Conservative 0; Mismatches 270; Indels 10; Gaps 2;

[illegible]

RESULT 15
A01100502

[illegible]

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1. Kiebauduaue; Perouglinae; Caenolabidulls.
1 (bases 1 to 345)
kohara,y., Shin-i,t., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
kohara,y., Shin-i,t., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
A complementary view of the C. elegans genome
Unpublished (2000)
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan

Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: yokohara@lab.nig.ac.jp.

FEATURES

source
1..345
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk709a4"
/clone_lib="unpublished oligo-capped cDNA library"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"
86 c 67 g 91 t 1 others

BASE COUNT
ORIGIN

Query Match 3.6%; Score 50.4; DB 10; Length 345;
Best Local Similarity 69.0%; Pred. NO. 2;
Matches 69; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
Qy 52 ccaagagacaagtacaattgtctactgtcgctcattctgttggtgagtcggagttctt 111
Db 106 CCAGAGGACAGGGAATCTTGTTTCACATTATCTGCCATCGGAATCGGAACCTCTT 165
Qy 112 ctgccatggaatatgttcattactatcgccctcgagtatt 151
Db 166 ATGCCATGGACATGCTTATTACGATCTCCATGATTATT 205

Search completed: February 28, 2002, 00:25:49
Job time: 3830 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2002, 23:24:08 ; Search time 49.03 Seconds
(without alignments)
6411.411 Million cell updates/sec

Title: US-08-816-011f-36
Perfect score: 1388
Sequence: 1 atggtaataatcaacgacgc.....ttattataaaaaaaaaa 1388

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323899 residues
Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Missing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Description	
Result No.	Score	Query Match	Length DB ID		
1	68	4.9	7218 1	US-08-232-463-14	Sequence 14, Appl
2	38.6	2.8	1582 3	US-08-545-196B-10	Sequence 10, Appl
3	38.6	2.8	1582 3	US-08-545-196B-12	Sequence 12, Appl
4	36	2.6	1164 2	US-08-794-796-1	Sequence 1, Appl
5	36	2.6	2621 2	US-08-553-619B-8	Sequence 8, Appl
6	36	2.6	5852 1	US-07-867-106-2	Sequence 2, Appl
7	35.8	2.6	1400 1	US-08-464-164-1	Sequence 1, Appl
8	35.8	2.6	1400 1	US-08-338-057-1	Sequence 1, Appl
9	35.8	2.6	1400 2	US-08-668-416-1	Sequence 1, Appl
10	35.4	2.6	581 3	US-08-557-309B-22	Sequence 22, Appl
11	35.4	2.6	581 3	US-08-834-306-22	Sequence 22, Appl
12	35.4	2.6	581 4	US-08-993-674A-22	Sequence 22, Appl
13	35.4	2.6	1303 2	US-08-793-410-11	Sequence 11, Appl
14	35.4	2.6	3489 2	US-08-728-323A-1	Sequence 1, Appl
15	35.4	2.6	32207 2	US-08-770-379-20	Sequence 20, Appl
16	35.4	2.6	32207 4	US-08-757-669A-20	Sequence 20, Appl
17	35.2	2.5	972 3	US-09-286-690-1	Sequence 1, Appl
18	35	2.5	51259 3	US-08-781-891-209	Sequence 209, App
19	34.8	2.5	289 4	US-09-007-005-17	Sequence 17, Appl
20	34.8	2.5	289 4	US-09-244-796-17	Sequence 17, Appl
21	34.6	2.5	991 4	US-09-377-648-7	Sequence 7, Appl
22	34.6	2.5	2238 1	US-08-742-011-1	Sequence 1, Appl
23	34.6	2.5	3645 2	US-08-663-112-1	Sequence 1, Appl
24	34.2	2.5	852 4	US-09-461-697-1	Sequence 1, Appl
25	34.2	2.5	1991 2	US-08-415-593-40	Sequence 40, Appl
26	34.2	2.5	2128 2	US-08-415-593-39	Sequence 39, Appl
27	33.8	2.4	519 1	US-08-339-582-1	Sequence 1, Appl

28	33.6	2.4	732 4	US-08-916-576B-5	Sequence 5, Appli
29	33.4	2.4	756 2	US-08-530-165-1	Sequence 1, Appli
30	33.2	2.4	174 2	US-08-378-235B-8	Sequence 8, Appli
31	33.2	2.4	248 4	US-09-007-005-32	Sequence 32, Appl
32	33.2	2.4	248 4	US-09-244-796-32	Sequence 32, Appl
33	33.2	2.4	277 4	US-09-007-005-3	Sequence 3, Appli
34	33.2	2.4	277 4	US-09-244-796-3	Sequence 3, Appli
35	33.2	2.4	1364 1	US-08-265-087-3	Sequence 3, Appli
36	33.2	2.4	1364 1	US-08-621-493-3	Sequence 3, Appli
37	33.2	2.4	1364 2	US-08-965-688-3	Sequence 3, Appli
38	33.2	2.4	1364 4	US-09-260-173-3	Sequence 3, Appli
39	33.2	2.4	2265 3	US-09-369-618-3	Sequence 3, Appli
40	33.2	2.4	2265 3	US-09-369-617-3	Sequence 3, Appli
41	33	2.4	306 4	US-09-122-400B-9	Sequence 9, Appli
42	33	2.4	1939 1	US-07-715-751B-2	Sequence 2, Appli
43	33	2.4	3498 3	US-08-293-728-1	Sequence 1, Appli
44	33	2.4	3498 4	US-09-421-868-1	Sequence 1, Appli
45	33	2.4	5852 1	US-07-867-106-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

[illegible]

RESULT 2
US-08-545-196B-10
; Sequence 10, Application US/08545196B
; Patent No. 6080577
; GENERAL INFORMATION:
; APPLICANT: MELKI, JUDITH
; APPLICANT: MURNICH, ARNOLD
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
; TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,196B
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. J.
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 2121-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA

```

US-08-545-196B-10
Query Match 2.8%; Score 38.6; DB 3; Length 1582;
Best Local Similarity 63.4%; Pred. No. 0.17;
Matches 59; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 1296 tatctataaattattatagcattagctactgttctatgttttttatttaagctg 1355
      ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 1441 TATCTTCATATGTTTAAAGTATATAAATAATATTAAATTTTTTAAAAA 1500
      ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||

Qy 1356 tggataaataattattataaataaataa 1388
      || ||||| || || ||||| || ||||| || ||||| || ||||| ||
Db 1501 AAAAAA 1533

RESULT 3
US-08-545-196B-12
; Sequence 12, Application US/08545196B
; Patent No. 6080577
; GENERAL INFORMATION:
; APPLICANT: MELKI, JUDITH
; APPLICANT: MUNNICH, ARNOLD
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
; NUMBER OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,196B
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. J.
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 2121-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-545-196B-12

Query Match 2.8%; Score 38.6; DB 3; Length 1582;
Best Local Similarity 63.4%; Pred. No. 0.17;
Matches 59; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```

RESULT 4

COMPUTER RESEARCH LAB.

TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
NAME/KEY: CDS
LOCATION: 2378..5038
US-07-867-106-2

Query Match 2.6%; Score 36; DB 1; Length 5852;
Best Local Similarity 55.6%; Pred. No. 1.8;
Matches 69; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
QY 1265 ttgtattgagcacttcgtggacagccaggtatcttataataattttatagcattagagt 1324
Db 2096 TTATTAATAAAAAAATCAAAAAAACCAAGTAATATTTATATATGAGGGTTTTT 2037
QY 1325 atacttgattatgttttttattagctgtggataataataattattaaaaa 1384
Db 2036 TTTTTTTTTTTTTTTTTTTCAAGTAAAAAATAAAAAAAGAAATA 1977
QY 1385 aaaa 1388
Db 1976 GAAA 1973

RESULT 7
US-08-464-164-1
Sequence 1, Application US/08464164
Patent No. 5614195
GENERAL INFORMATION:
APPLICANT: Tomley, Fiona M.
APPLICANT: Dunn, Paul P. J.
APPLICANT: Bumstead, Janene M.
APPLICANT: Vermeulen, Arno N.
TITLE OF INVENTION: Coccidiosis poultry vaccine
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Akzo No. 5614195el Patent Department
STREET: 1300 Piccard Drive, Suite 206
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,164
FILING DATE: June 2, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gormley, Mary E.
REGISTRATION NUMBER: 34,409
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1400 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
ORIGINAL SOURCE:
ORGANISM: Eimeria maxima
STRAIN: Houghton
DEVELOPMENTAL STAGE: sporozoite
IMMEDIATE SOURCE:
LIBRARY: sporozoite cDNA cloned in Lambda ZAPII
CLONE: Em70-1
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1368
US-08-464-164-1

Query Match 2.6%; Score 35.8; DB 1; Length 1400;
Best Local Similarity 69.0%; Pred. No. 1;
Matches 49; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 1318 tttagtgataactgttatgtttttttattagctgtggataataataattattaaaa 1377
Db 1326 TCAGCAAAATGCTTCTTAATTTATGTGTAATCTGCAGCAGATAATAATAATAATAA 1385
QY 1378 aaaaaaaaaa 1388
Db 1386 AAAAAAAAAA 1396

RESULT 8
US-08-338-057-1
Sequence 1, Application US/08338057
Patent No. 5795741
GENERAL INFORMATION:
APPLICANT: Tomley, Fiona M.
APPLICANT: Dunn, Paul P. J.
APPLICANT: Bumstead, Janene M.
APPLICANT: Vermeulen, Arno N.
TITLE OF INVENTION: Coccidiosis poultry vaccine
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
STREET: 1330 Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/338,057
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 93.309078.9
FILING DATE: 12-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gormley, Mary E.
REGISTRATION NUMBER: 34,409
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1400 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

Query Match 2.6%; Score 35.8; DB 2; Length 1400;
Best Local Similarity 69.0%; Pred. No. 1;
Matches 49; Conservative 0; Mismatches 22; Indels 0

[illegible]

RESULT 10

```

: Sequence 22, Application US/08557309B
: Patent No. 5916572
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Lodes, Michael J.
: APPLICANT: HoughCon, Raymond L.
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TH
: NUMBER OF SEQUENCES: 69
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED AND BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: USA
: ZIP: 98104-7092
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/557,309B
: FILING DATE: 14-NOV-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Makl, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 210121.422
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 22:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 581 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-08-557-309B-22

```

Query Match 2.6%; Score 35.4; DB 2; Length 581;
Best Local Similarity 61.3%; Pred. No. 0.83;

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs
TYPE: nucleic acid

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

Query Match 2.6%; Score 35.4; DB 2; Length 32207;
Best Local Similarity 57.8%; Pred. No. 6.3;
Matches 63; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
Qy 264 cctgttctcattattgctggtccctgatctaccgcggtctttgtctcgggttgcttcaa 323
Db 20715 CCTCTCGTCTCTCTTGTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAT 20774
Qy 324 catcgtcaacctgacaatcatctcctcgtcgtcattgttctgagccc 372
Db 20775 CATGTCATCCTC 20823

Search completed: February 28, 2002, 00:50:28
Job time: 5180 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 16:59:34 ; Search time 13.23 seconds
(without alignments)
1202.762 Million cell updates/sec

Title: US-08-816-011f-63
Perfect score: 2294
Sequence: 1 MVIINRSNTYAVQEAEPD.....LTGGLMPVIEHFVDRKPSIL 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	343.5	15.0	455	1	ENT1_HUMAN	Q99808	homo sapien
2	326.5	14.2	456	1	ENT2_HUMAN	Q14542	homo sapien
3	320.5	14.0	456	1	ENT2_RAT	O54699	rattus norv
4	301	13.1	456	1	ENT1_RAT	O54699	rattus norv
5	209.5	9.1	327	1	ENT2_MOUSE	Q61672	m equibila
6	182.5	8.0	517	1	FU26_YEAST	P31381	saccharomyc
7	134	5.8	432	1	Y672_METUA	Q58086	methanococc
8	124.5	5.4	506	1	HTFB_HAEIN	P71338	haemophilus
9	117.5	5.1	559	1	YABC_SCHPO	Q09852	schizosacch
10	113.5	4.9	477	1	Y098_MYCGE	P47344	mycoplasma
11	113.5	4.9	587	1	T953_MOUSE	Q9et30	mus musculus
12	113.5	4.9	589	1	T953_HUMAN	Q9rd45	homo sapien
13	113.5	4.9	604	1	Y09M_HORSE	P14656	equus caball
14	113	4.9	459	1	TCR2_BACSU	P14512	bacillus su
15	113	4.9	614	1	YDNK_LACLC	P42377	lactococcus
16	112	4.9	383	1	NAPA_ENTHR	P26235	enterococcu
17	111.5	4.9	397	1	PAR2_HUMAN	P55085	homo sapien
18	111.5	4.9	433	1	TCR_STAUI	P02983	staphylococ
19	111.5	4.9	439	1	UHPG_ECOLI	P09836	escherichia
20	110.5	4.8	479	1	Y098_MYCPN	P75535	mycoplasma
21	110.5	4.8	785	1	ISP4_SCHPO	P40900	schizosacch
22	108.5	4.7	442	1	UHPG_SALTY	P27669	salmonella
23	108	4.7	388	1	YUBA_BACSU	Q32086	bacillus su
24	106.5	4.6	399	1	PAR2_MOUSE	P55086	mus musculus
25	106	4.6	547	1	N5M_ASCSU	P24884	ascaris suu
26	105.5	4.6	598	1	THIX_YEAST	Q08485	saccharomyc
27	105	4.6	542	1	ATRI_YEAST	P13090	saccharomyc
28	104.5	4.6	894	1	YN86_YEAST	P27514	saccharomyc
29	104	4.5	527	1	NU2M_ACACA	Q37376	acanthamoeb
30	104	4.5	552	1	N5M_RHISA	Q92ym7	rhinicephal
31	103.5	4.5	345	1	NUOH_RHOCA	P42032	rhodobacter
32	103.5	4.5	606	1	N5M_EQUAS	P92485	equus asinu
33	103	4.5	459	1	NU4M_BOVIN	P03910	bos taurus

RESULT 1

ID	ENT1_HUMAN	STANDARD;	PRT;	455 AA.
AC	Q99808;			
DT	20-AUG-2001 (Rel. 40, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	EQUILIBRATIVE NUCLEOSIDE TRANSPORTER 1 (EQUILIBRATIVE			
DE	NITROBENZYLMECAPTOPURINE RIBOSIDE-SENSITIVE NUCLEOSIDE TRANSPORTER)			
DE	(EQUILIBRATIVE NMPR-SENSITIVE NUCLEOSIDE TRANSPORTER) (NUCLEOSIDE			
DE	TRANSPORTER, ES-TYPE).			
GN	SLC29A1 OR ENTI.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	[1]			
RN	SEQUENCE FROM N.A., AND SEQUENCE OF 1-21.			
RC	TISSUE=Placenta;			
RX	MEDLINE=97140266; PubMed=8986748;			
RA	Griffiths M., Beaumont N., Yao S.Y.M., Sundaram M., Boumah C.E.,			
RA	Davies A., Kwong F.Y.P., Coe I., Cass C.E., Young J.D., Baldwin S.A.;			
RT	"Cloning of a human nucleoside transporter implicated in the cellular			
RT	uptake of adenosine and chemotherapeutic drugs.;"			
RL	Nat. Med. 3:89-93(1997).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Small intestine, and Jejunum;			
RA	Lum P.Y., Ngo L.Y., Bakken A.H., Unadkat J.D.;			
RT	"Critical structural determinants for high affinity binding of			
RT	nucleosides to the equilibrative NMPR-sensitive nucleoside			
RT	transporter (es) cloned from the human jejunum.;"			
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.			
CC	!- FUNCTION: MEDIATES BOTH INFUX AND EFFLUX OF NUCLEOSIDES ACROSS			
CC	THE MEMBRANE (EQUILIBRATIVE TRANSPORTER). IT IS SENSITIVE (ES) TO			
CC	LOW CONCENTRATIONS OF THE INHIBITOR NITROBENZYLMECAPTOPURINE			
CC	RIBOSIDE (NMPR) AND IS SODIUM-INDEPENDENT. IT HAS A HIGHER			
CC	AFFINITY FOR ADENOSINE. INHIBITED BY DIPYRIDAMOLE AND DILAZEP			
CC	(ANTICANCER CHEMOTHERAPEUTICS DRUGS).			
CC	!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.			
CC	!- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, MAMMARY GLAND,			
CC	ERYTHROCYTES AND PLACENTA, AND ALSO IN FETAL LIVER AND SPLEEN.			
CC	!- PTM: GLYCOSYLATED.			
CC	!- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; U81375; AAC51103.1; -			
DR	EMBL; AF079117; AAC62495.1; -			
DR	MIM; 602193; -			

051750 borrelia bu
037370 acanthamoeb
033814 staphylococ
060779 homo sapien
046522 b cytochrom
P49219 xenopus lae
P39637 bacillus su
P39301 escherichia
P04371 trypanosoma
P57543 buchnera ap
Q46916 escherichia
P25568 saccharomyc

ALIGNMENTS


```
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 124 144 POTENTIAL.
FT TRANSMEM 162 182 POTENTIAL.
FT TRANSMEM 193 213 POTENTIAL.
FT TRANSMEM 291 311 POTENTIAL.
FT TRANSMEM 324 344 POTENTIAL.
FT TRANSMEM 360 380 POTENTIAL.
FT TRANSMEM 386 406 POTENTIAL.
FT TRANSMEM 432 452 POTENTIAL.
FT CARBOHYD 48 48 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 130 MISSING (IN ISOFORM HNP36).
FT CONFLICT 200 200 Y -> C (IN REF. 1).
SQ SEQUENCE 456 AA; 50172 MW; AEF1B244397508E1 CRC64;

Query Match 14.2%; Score 326.5; DB 1; Length 456;
Best Local Similarity 22.9%; Pred. No. 1.le-15;
Matches 106; Conservative 71; Mismatches 200; Indels 85; Gaps 11;

Qy 18 PRDKYNIYVWLVILVCGVLLPWNMFITIAPEY-----VNYWFKPDGVE 62
||| ||| :|| :||| ||| | :| :|
Db 7 PRDSVHLVGISFFILGLTGLLPPWNFFITAIYFORLAGACNSTARILSTNHTGPEDAF- 65
||| ||| :||| :||| ||| | :| :|
Qy 63 TWYSEKFGSLTIGSQLPNASINVENFLIITAGPLIYRVFAPVCFNIVNLITILILVIVL 122
||| ||| :||| :||| ||| | :| :|
Db 66 -----NENWYVLLSOLPLLLFTLLNSFLYQCVPETVRILG-----SLAIIILLFALTRAL 116
||| ||| :||| :||| ||| | :| :|
Qy 123 EPTDSMSWFFWTLTGMTSINFSNGLYVNSVYGVGGDFPHYTGALLIGNNICGLLITV 182
||| ||| :||| :||| ||| | :| :|
Db 117 VKVDSMPGPFISWASVCFINSFSAVLQGLSLGQLTGMPSTVTLFSLGQGLAGIFAAL 176
||| ||| :||| :||| ||| | :| :|
Qy 183 VKI-----GVYFLNDEPKLVAIVF---GISLVILLVCAITAFITTKODFY---HYHHQ 231
||| ||| :||| :||| ||| | :| :|
Db 177 AMLLSMASGV-----DAETSALGYFITPYVGIILMSIVCYLSPLHLKFARYLYLANKSSQA 230
||| ||| :||| :||| ||| | :| :|
Qy 232 KGEIREKA-----ETDRPSPSLTWTTFNC 257
||| ||| :||| :||| ||| | :| :|
Db 231 QAOLETRAELLOSDENGIPSSPQKVALTLDLLEKEPESEDEPKCKPSVFTVFOKI 290
||| ||| :||| :||| ||| | :| :|
Qy 258 YQLENVWFCAVTTIPTVMVTVTRGSGPLNKIMSENDEIYLLSFLVFNFAIG 317
||| ||| :||| :||| ||| | :| :|
Db 291 WLTAICLVLTFTLSVFPATAMVTSSTS-----PGKWSOFFNPICCFLLFNIMDMWLG 344
||| ||| :||| :||| ||| | :| :|
Qy 318 SVASKIHP--TPYLKFAILLRALFTPEFFFCNVRQTRAYPVEESTDLFVIGGAM 375
||| ||| :||| :||| ||| | :| :|
Db 345 RSLTSYFLWPDBDSRLPLLLVLCRLFLVPLFMLCHVPQNSR-LPIFLPQDAYFITFMLLF 403
||| ||| :||| :||| ||| | :| :|
Qy 376 SFSGHLSALAMGYTPNVVPSHYSFRAQLSVCTLMVGLLTG 417
||| ||| :||| :||| ||| | :| :|
Db 404 AVSNGYLVSLTMCLAPROVLPHEREVAGALMTFFLALGLSCG 445
||| ||| :||| :||| ||| | :| :|

RESULT 3
ENT2_RAT STANDARD; PRT; 456 AA.
AC O54699;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EQUILIBRATIVE NUCLEOSIDE TRANSPORTER 2 (EQUILIBRATIVE
DE NITROBENZYLMECAPTOPURINE RIBOSIDE-INSENSITIVE NUCLEOSIDE TRANSPORTER)
DE (EQUILIBRATIVE NEMPR-INSENSITIVE NUCLEOSIDE TRANSPORTER) (NUCLEOSIDE
DE TRANSPORTER, EI-TYPE).
GN SLC29A2 OR ENT2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Jejunum;
```

```
RX MEDLINE=98019212; PubMed=9353301;
RA Yao S.Y.M., Ng A.M.L., Muzyka W.R., Griffiths M., Cass C.E.,
RA Baldwin S.A., Young J.D.;
RT "Molecular cloning and functional characterization of
RT nitrobenzylthioinosine (NBMPR)-sensitive (es) and NBMPR-insensitive
RT (ei) equilibrative nucleoside transporter proteins (rENT1 and rENT2)
RT from rat tissues."
RL J. Biol. Chem. 272:28423-28430(1997).
CC -!- FUNCTION: MEDIATES BOTH INFLUX AND EFFLUX OF NUCLEOSIDES ACROSS
CC THE MEMBRANE (EQUILIBRATIVE TRANSPORTER). IT IS INSENSITIVE (EI)
CC TO LOW CONCENTRATIONS OF THE INHIBITOR NITROBENZYLMECAPTOPURINE
CC RIBOSIDE (NBMPR) AND IS SODIUM-INDEPENDENT. SPECIFIC FOR
CC NUCLEOSIDES, BUT MAY ALSO TRANSPORT HYPOXANTHINE. MAY ALSO PLAY A
CC ROLE IN THE EFFLUX OF INOSINE AND HYPOXANTHINE FROM MUSCLE CELLS
CC DURING THE NET DEGRADATION OF PURINE NUCLEOTIDES THAT OCCURS
CC DURING STRENUOUS EXERCISE AND/OR IN THE REUPTAKE OF THESE PURINES
CC DURING THE RECOVERY PROCESS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- MISCELLANEOUS: RESISTANT TO DIPYRIDAMOLE AND DILAZEP INHIBITION
CC (ANTICANCER CHEMOTHERAPEUTICS DRUGS).
CC -!- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF015305; AAB88050.1;
DR InterPro; IPR002259; DER_Nucleoside_tran.
DR Pfam; PF01733; Nucleoside_tran; 1.
DR PRINTS; PR01130; DERENTRNSPRT.
DR ProDom; PD005103; DER_Nucleoside_tran; 1.
KW Transmembrane; Transport; Glycoprotein.
FT DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 13 29 POTENTIAL.
FT DOMAIN 30 68 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 69 93 POTENTIAL.
FT DOMAIN 94 97 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 98 116 POTENTIAL.
FT DOMAIN 117 124 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 125 143 POTENTIAL.
FT DOMAIN 144 160 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 161 185 POTENTIAL.
FT DOMAIN 186 192 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 193 213 POTENTIAL.
FT DOMAIN 214 291 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 292 311 POTENTIAL.
FT DOMAIN 312 323 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 324 342 POTENTIAL.
FT DOMAIN 343 359 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 360 378 POTENTIAL.
FT DOMAIN 379 393 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 394 413 POTENTIAL.
FT DOMAIN 414 431 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 432 452 POTENTIAL.
FT DOMAIN 453 456 EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 456 AA; 50265 MW; DA97C2C578E1EE9D CRC64;

Query Match 14.0%; Score 320.5; DB 1; Length 456;
Best Local Similarity 23.5%; Pred. No. 2.8e-15;
Matches 107; Conservative 76; Mismatches 200; Indels 73; Gaps 13;

Qy 18 PRDKYNIYVWLVILVCGVLLPWNMFITIAPEYVNYWFKPDGVEYWSK-----EF 69
||| ||| :||| :||| ||| | :| :|
Db 7 PRDSVHLVGISFFILGLTGLLPPWNFFITAIYFORLAGACNSTARILSTNHTGPEDAF 66
||| ||| :||| :||| ||| | :| :|
Qy 70 MGSLTIGSQLPNASINVENFLIITAGPLIYRVFAPVCFNIVNLITILILVILEPTEDSM 129
```

Db 67 NNWYTLSSQLPPLLFTLLNSFLYQCIPESVRILG-----SLIAILLFALTAALVKVDLSP 122
Qy 130 SWRFWVTLGMAATSNFNGLYENSVYGVGGDFPHYTGALLIGNNICGL-----LIIVVK 184
Db 123 GLFESITMASWVFNSCAVLGSLFQOLGMPSTYSTLFLSGGGLAGIFAALAMLTSLA 182
Qy 185 IGVTFYFLNDPKLVAIVYF-----GISLVILVCAIAL-----FFITKQ----- 223
Db 183 SGV-----DPQTSALGYFITPCVGILLISICYLSPHLKFAFYLLTKKQAPVQLELETK 236
Qy 224 -DFVHYHQKMEIR-----EK-----AETDRSPSILMTTTCYVGLFNV 264
Db 237 AELLGADEKNGIPVSPQOAGPTLDLDPKELEGLEEPPQKPGKPSVFWFKIWLTLCL 296
Qy 265 WFCFAVTLTIP-VMMVTVTTRGDSGLFNKIMSENDEIYTLTSLFVNLFAAGISVASK 323
Db 297 VLVTFTVLSVFPATAMVTSSNS-----PGKWSQFFNPICCLLFNVMDWGLRSITSY 350
Qy 324 IHWP--TPRYLKFAIILRALFIPFFFCNRYVQTRAYPVFFESTDIFVIGIAMSFSGHY 381
Db 351 FLWPDSDQLPLLVCLRFLEVPFLMCHVPQAR-LPIIFWQDAYFITPMLLFAISNGY 409
Qy 382 LSLAMGYTPNVVPSHYRFAAQLSVCTLMVGLLTG 417
Db 410 FVSLTMCAPRQVLPHEREVAGALMTFFLALGLSCG 445

RESULT 4

ENTL_RAT ID ENTL_RAT STANDARD; PRT; 456 AA.
AC 054698;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EQUILIBRATIVE NUCLEOSIDE TRANSPORTER 1 (EQUILIBRATIVE
DE NITROBENZYLMECAPTOPURINE RIBOSIDE-SENSITIVE NUCLEOSIDE TRANSPORTER)
DE (EQUILIBRATIVE NEMPR-SENSITIVE NUCLEOSIDE TRANSPORTER) (NUCLEOSIDE
DE TRANSPORTER, ES-TYPE).
GN SLC29A1 OR ENTL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Jejunum;
RX MEDLINE=98019212; PubMed=933301;
RA Yao S.Y.M., Ng A.M.L., Muzyka W.R., Griffiths M., Cass C.E.,
RA Baldwin S.A., Young J.D.;
RT "Molecular cloning and functional characterization of
RT nitrobenzylchinosine (NEMPR)-sensitive (es) and NEMPR-insensitive
RT (ei) equilibrative nucleoside transporter proteins (rent1 and rent2)
RT from rat tissues";
RL J. Biol. Chem. 272:28423-28430(1997).
CC -!- FUNCTION: MEDIATES BOTH INFLUX AND EFFLUX OF NUCLEOSIDES ACROSS
CC THE MEMBRANE (EQUILIBRATIVE TRANSPORTER). IT IS SENSITIVE (ES) TO
CC LOW CONCENTRATIONS OF THE INHIBITOR NITROBENZYLMECAPTOPURINE
CC RIBOSIDE (NEMPR) AND IS SODIUM-INDEPENDENT. IT HAS A HIGHER
CC AFFINITY FOR ADENOSINE, RESISTANT TO DIPHENIDOL AND DILAZEP
CC INHIBITION (ANTICANCER CHEMOTHERAPEUTICS DRUGS).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN JEJENUM, LIVER AND LUNG.
CC -!- PTM: GLYCOSYLATED.
CC -!- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).

CC EMBL; AF015304; AAB88049.1; --
DR InterPro; IPR002259; DER_Nucleoside_tran.
DR Pfam; PF01733; Nucleoside_tran; 2;
DR PRINTS; PR01130; DERENTRNSPRT.
DR ProDom; PD005103; DER_Nucleoside_tran; 1.
KW Transmembrane; Transport; Glycoprotein.
FT INIT_MET 0 0
FT DOMAIN 1 11 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 12 28 POTENTIAL.
FT DOMAIN 29 81 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 82 106 POTENTIAL.
FT DOMAIN 107 110 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 111 129 POTENTIAL.
FT DOMAIN 130 137 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 138 156 POTENTIAL.
FT DOMAIN 157 173 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 174 198 POTENTIAL.
FT DOMAIN 199 205 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 206 226 POTENTIAL.
FT DOMAIN 227 290 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 291 310 POTENTIAL.
FT DOMAIN 311 322 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 323 342 POTENTIAL.
FT DOMAIN 343 359 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 360 378 POTENTIAL.
FT DOMAIN 379 393 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 394 413 POTENTIAL.
FT DOMAIN 414 431 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 432 452 POTENTIAL.
FT DOMAIN 453 456 EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 456 AA; 49885 MW; A34CE92C20836D9B CRC64;

Query Match. 13.1%; Score 301; DB 1; Length 456;

Best Local Similarity 22.3%; Pred. No. 6.2e-14;

Matches 107; Conservative 84; Mismatches 191; Indels 98; Gaps 17;

Qy 18 PRKYNIVYWLVLVGVGLLPNNMTTIAPEYVYVWFKPDGV-----ETWYSKEPMGS 72
Db 6 PDQRYKAVWLIFVGLGLTLLPNWFFIT-ATQYFTRSLNTSQNISLVTSNCSCESTALAD 64
Qy 73 LTIGSQLPNASINVFNLFIIAGPLIYRVAPVCFN-----IVNLTII 115
Db 65 PSVSLPARSSLSAIFNNVWTLCAMLPLLIIF--TCLNLFHQKVSQSRLILGSLALLVF 122
Qy 116 LILVILVLEPTEDSMWFFWVTLGMAATSNFNGLYENSVYGVGGDFPHYTGALLIGNNI 175
Db 123 LVTATLVKQVMDALS-PFIITMIKIVLINSFGAILQASLFLAGVLPANVTAPIMSGQL 181
Qy 176 CGLLITVVKI-----GVTFPLNDPKLVAIVYFGLSVLILLVCAIAFFITKQ 223
Db 182 AGFTSVAMICAVASGSKLSAFAFYFIT---ACAVV-----ILAICYLALPWM--- 228
Qy 224 DFVYHHQ-----KMEIR-EKAETDRSPSL----- 250
Db 229 EYRHYQLQLNLAGPAOETKLDLISEGEEPRGGREESGVPGPNLSPANRQSTKAIUKSI 288
Qy 251 WTTFTTCYGLFNWFCFAVTLTIFPVMTVTTRGDSGLFNKIMSENDEIYTLTSLFV 310
Db 289 WVL-----ALSCVFIFVTIGLFP---AVTAEVSSIAGTSPWKNK-FEIPVACFLNF 337
Qy 311 NLFAAIGSVASKIHP--TPRYLKFAIILRALFIPFFFCNRYVQTRAYPVFFESTDIF 368
Db 338 NVFDWGLRSLSLTAICMPPGQDSRWLPVLVACRVVFIPILLMLCNVK-QHHYLPSPFKHDWF 396
Qy 369 VIGGIAMSFSGHVSALAMGYTPNVVPSHYRFAAQLSVCTLMVGLLTGGLWVPVIEHF 428
Db 397 ITFMAAFASNGYLSLCLMCFGPKVKPAPAEYAGNIMSFLLCLGLALGAVLSFLRLV 456

```

RESULT 5
ENT2_MOUSE
ID ENT2_MOUSE STANDARD; PRT; 327 AA.
AC Q61672;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EQUILIBRATIVE NUCLEOSIDE TRANSPORTER 2 (EQUILIBRATIVE
DE NITROBENZYL MERCAPTOPYRINE RIBOSIDE-INSENSITIVE NUCLEOSIDE TRANSPORTER)
DE (EQUILIBRATIVE NMMP-INSENSITIVE NUCLEOSIDE TRANSPORTER) (NUCLEOSIDE
DE TRANSPORTER, EI-TYPE) (36 KDA NUCLEOLAR PROTEIN HNP36) (HYDROPHOBIC
DE NUCLEOLAR PROTEIN, 36 KDA) (DELAYED-EARLY RESPONSE PROTEIN 12).
GN SLC29A2 OR ENT2 OR HNP36 OR DER12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=BALB/C; TISSUE=Fibroblast;
RC MEDLINE=95367016; PubMed=7639753;
RA Williams J.B., Lanahan A.A.;
RT "A mammalian delayed-early response gene encodes HNP36, a novel,
RT conserved nucleolar protein.";
RL Biochem. Biophys. Res. Commun. 213:325-333(1995).
CC -1- FUNCTION: MEDIATES EQUILIBRATIVE TRANSPORT OF PURINE AND
CC PYRIMIDINE NUCLEOSIDES, AND THE PURINE BASE HYPOXANTHINE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLAR. INTEGRAL MEMBRANE
CC PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM AND A SHORT
CC FORM/HNP36 (SHOWN HERE); SEEMS TO BE PRODUCED BY ALTERNATIVE
CC SPLICING.
CC -1- INDUCTION: BY PLATELET DERIVED GROWTH FACTOR (PDGF) AND FIBROBLAST
CC GROWTH FACTOR (FGF).
CC -1- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X86682; CAA60381.1; -
CC MGP; MGI:1345278; SLC29a2.
DR InterPro: IPR002259; DER_Nucleoside_tran.
DR Pfam: PF01733; Nucleoside_tran; 1.
DR PRINTS; PR01130; DERENTRNSPT.
DR ProDom; PD005103; DERENTRNSPT.
KW Nuclear protein; Transmembrane; Transport; Alternative splicing.
FT TRANSMEM 2 22
FT TRANSMEM 32 52
FT TRANSMEM 63 83
FT TRANSMEM 159 179
FT TRANSMEM 194 214
FT TRANSMEM 231 251
FT TRANSMEM 267 287
FT TRANSMEM 303 323
SQ SEQUENCE 327 AA; 36113 MW; 5D2D3FF4BBD592B6 CRC64;

Query Match 9.1%; Score 209.5; DB 1; Length 327;
Best Local Similarity 22.4%; Pred. No. 8.5e-08;
Matches 76; Conservative 49; Mismatches 134; Indels 81; Gaps 12;

QY 131 WFFVFWTLGMATSIINFSNGLYNSVYGVGDFPHPTVIGALLICNNICGLLITVWKI----- 185
|| : : : : : | | | | | : : : : :
Dd 5 WF-----INSFCAVGLSFLGQLGTMPSTYTLFSLGGLAGIFAALMLMSLAS 54
|| : : : : : | | | | | : : : : :
QY 186 GVTFLNDPKLVAIVF---GISIVLLVCAIAL-----FFI-----TK 222
|| : : : : : | | | | | : : : : :

```

```
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 151 171 POTENTIAL.
FT TRANSMEM 174 194 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 243 263 POTENTIAL.
FT TRANSMEM 344 364 POTENTIAL.
FT TRANSMEM 367 387 POTENTIAL.
FT TRANSMEM 411 431 POTENTIAL.
FT TRANSMEM 446 466 POTENTIAL.
FT TRANSMEM 492 512 POTENTIAL.
SQ SEQUENCE 517 AA; 43C92A3E9A3D8D50 CRC64;

Query Match 8.0%; Score 182.5; DB 1; Length 517;
Best Local Similarity 19.9%; Pred. No. 9.4e-06;
Matches 79; Conservative 69; Mismatches 131; Indels 117; Gaps 20;

QY 23 NIVYWLIVLGVGLLWNNFIITAPIYVYVWFKPGCVETW----YSKEFGMSLTIGSQL 79
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 75 NLSYITFAIGIGLLWPNCILS-ASQYFKHDIFKDTSI--WAKIFTSSMMSESTISSML 131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 80 PNASINVENLFL-----IIAGPLIYR--VFAPVCENIVNLTILILIVIVLEPTED 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 132 -----FNIYLAQRQYKRYRRVING-LVWEIIVFTVMCF-----FTILHFL----- 171

QY 128 SMSWFEWTLGMAITSI-NFSGNLYENSVYGGDFPHYTCALLIGNNICGLLITVVKIG 186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 -PKWFNEMFIMLWVSSMTAMTQNGIMAIANVFGSEYSGQVMVGAVGLPSLVIFA 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 187 VTVELNDEPKLVA--IVYFGISLVILLVCAIALFFITK----- 222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 LAFTIENSVSTTGILLYFFTTLLVTVTIC-VVMFVSVKISKRVNENWVEDGHITDVLG 289

QY 223 -----QDFYVH-----HOKGMEIREKAETDRPSPSILWTTFTNC 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 SLRSNEEIRIVGRIDQMEDHRTNGTRDDNDEGEELQKVPFE-----VLFACL 341

QY 258 YGOLFNVWFCFVTLTTPVWMTVTRGDSGLNKMSENDEIYTLTSLFVNLFNFAIG 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 342 KYLVLSITFTFVVTL-VFPVFASAT-----YVTGLPLSNAQVIPLI--FTLNLGLDLYG 392

QY 318 SIYASKIHWP-----TPRYLKAFAILLRALFIPFF 345
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 393 RVIAD---WPFNRDQKFTPRKTFIYSLRVAAPLPLF 425

RESULT 7
Y672_METJA
ID Y672_METJA STANDARD; PRT; 432 AA.
AC Q58086;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ0672.
GN MJ0672.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
   jannaschii.";
```


HITB_HAEBIN STANDARD; PRT; 506 AA.
 ID AC P71338; Q53440;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE IRON(III)-TRANSPORT SYSTEM PERMEASE PROTEIN HITB.
 GN HITB OR HI0096.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Keiley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.";
 RN Science 269:496-512(1995).
 RP SEQUENCE FROM N.A.
 RC STRAIN=NTHI TN106;
 RX MEDLINE=95012644; PubMed=7927717;
 RA Sanders J.D., Cope L.D., Hansen E.J.;
 RT Identification of a locus involved in the utilization of iron by
 RT Haemophilus influenzae";
 RN Infect. Immun. 62:4535-4525(1994).
 CC -!- FUNCTION: INVOLVED IN A PERIPLASMIC BINDING-PROTEIN-DEPENDENT
 CC IRON(III) TRANSPORT SYSTEM.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (POTENTIAL).
 CC -!- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
 CC PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE CYSTW
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U32695; AAC21774.1; ALT_INIT.
 CC EMBL; S72674; AAB32111.1; -.
 CC TIGR; HI0096; -.
 CC InterPro: IPR000515; BPD.transp.
 CC Pfam: PF00528; BPD.transp. 2.
 CC PROSITE: PS00402; BPD_TRANSP_INN_MEMBER; FALSE_NEG.
 CC Transport; Iron transport; Transmembrane; Inner membrane;
 KW Complete proteome.
 KW TRANSMEM 9 29 POTENTIAL.
 KW TRANSMEM 57 77 POTENTIAL.
 KW TRANSMEM 90 110 POTENTIAL.
 KW TRANSMEM 116 136 POTENTIAL.
 KW TRANSMEM 174 194 POTENTIAL.
 KW TRANSMEM 218 238 POTENTIAL.
 KW TRANSMEM 275 295 POTENTIAL.
 KW TRANSMEM 314 334 POTENTIAL.
 KW TRANSMEM 350 370 POTENTIAL.
 KW TRANSMEM 379 399 POTENTIAL.
 KW TRANSMEM 428 448 POTENTIAL.
 KW TRANSMEM 480 500 POTENTIAL.
 FT VARIANT 72 72 F -> L (IN STRAIN TN106).
 FT FT


```

RESULT 15
YDNK_LACLC
ID YDNK_LACLC STANDARD; PRT; 614 AA.
AC P42377;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 70.0 KDA PROTEIN IN DNAAK 3'REGION (ORF4).
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MG1363;
RX MEDLINE=94172317; PubMed=8126443;
RA Eaton T.J., Shearman C.A., Gasson M.J.;
RT "Cloning and sequence analysis of the dnaK gene region of Lactococcus
RT lactis subsp. lactis";
RL J. Gen. Microbiol. 139:3253-3263(1993).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X76642; CAA54090.1; --
CC Hypothetical protein.
KW SEQUENCE 614 AA; 69869 MW; B1146136D09B89D9 CRC64;
SQ
Query Match 4.9%; Score 113; DB 1; Length 614;
Best Local Similarity 19.3%; Pred. No. 0.65;
Matches 98; Conservative 66; Mismatches 149; Indels 196; Gaps 26;

Qy 21 KYNTVYWLVLVGVGLLPNNFTTIA---PEYVYVY----- 54
||| :||| ||| | :
Db 16 KYNEFFVHIIII-----FTTILATFPFYRGNFHAGNDFAFNRYARVNSTISALKDG 64

```

```

QY 55 ----WFKPDGVE----TWYKEPFMGSLTIGSQLPNASINVENFLIIITAGPLIYRVFAPVC 106
Db 65 QVTPQDPNALSFGYAW--NEEYV-----PLPTYFISVIK-FIVKSWLSLSLFSYSLC 115
QY 107 FNIIVNLTIILILVIVLEPTEDSMWFWVTLGMATSFNSGLYENSIVYGVGGDFPHYI 166
Db 116 LFTSGIFIFNFSSFLLDKHTNS-KLEGLLAVALT---FSNSTYINLYYANPSQPLALL 171
QY 167 GALLI--GNN-----ICG-----LLITVVKIGVT 188
Db 172 FVILLFWGMKMKNKRSFAAFMLVAFGAAGLPLSHVTTCITLPFVLLYLFLIIRKG-- 229
QY 189 YFLNDEKLVAIYVFGISLVILLVCAIALEF-----ITKQDFYH----- 227
Db 230 -NKENIKIIGLGLSVTSAI---GLSAFFLPLENLKSGLIYVNSNSDFSRSFGWNNI 284
QY 228 -YHQKGMETREKAETDRSPSILWTF-----TNC-YGQLFNWVFCFAVT 271
Db 285 AYFOGKWEPLYKIEFSYKFPSSLFVILFIFISLINEKKTNAKYSLIFS---CFSLV 341
QY 272 LTIPFVMMVTTRGDSGLFNKIMSENDEIYTLTTS---FLVENLFAAIG-----SIVAS 322
Db 342 LVLMLPI-----FPWKLES---IFTIVQDPARFSTLGLFSALSLLVILPILLD 388
QY 323 KIHWPTRYLKFAIILRALPIPFPPFCNRYRQTRAYPVF-----FESTDIF 368
Db 389 KISGKTSYLTIGLLVIFSLGFAEFNR-RIOKGSQPLFASQSLNKNKTPFNMYMENPDSI 447
QY 369 VIGGIAMSFHGYLSALAMGYTPNVVPSH 397
Db 448 AIG-----EYLPQVIGSH 460

```

Search completed: February 27, 2002, 17:02:49
Job time: 195 sec

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	2265	98.7	434	2	T16080	hypothetical prote
2	716.5	31.2	461	2	T28006	hypothetical prote
3	706.5	30.8	450	2	T23528	hypothetical prote
4	574.5	25.0	800	2	T23247	hypothetical prote
5	313.5	13.7	428	2	T01489	hypothetical prote
6	249.5	10.9	408	2	C86156	hypothetical prote
7	214	9.3	326	2	JC4196	36K hydrophobic nu
8	209.5	9.1	327	2	JC4195	36K hydrophobic nu
9	196.5	8.6	513	2	T21887	hypothetical prote
10	182.5	8.0	517	2	S36712	FUN26 protein - ye
11	180	7.8	143	2	T22164	hypothetical prote
12	144	6.3	418	2	D85064	hypothetical prote
13	138.5	6.0	418	2	E85064	hypothetical prote
14	134	5.8	432	2	H64383	Na+ transporter -
15	128.5	5.6	382	2	G96641	hypothetical prote
16	126	5.5	418	2	F83986	transporter BH2694
17	125.5	5.5	674	2	T21217	hypothetical prote
18	117.5	5.1	506	1	D64048	iron (III) ABC tra
19	117.5	5.1	559	2	S62503	inorganic phosphat
20	115	5.0	461	2	B83601	probable transport
21	114.5	5.0	449	2	C75053	DNA damage-inducib
22	113.5	4.9	477	2	H64210	eggshell protein p
23	113.5	4.9	604	2	T11867	NADH dehydrogenase
24	113	4.9	459	2	S42338	tetracyclin resist
25	113	4.9	614	2	S40085	hypothetical prote
26	112.5	4.9	694	2	E69143	hypothetical prote
27	112	4.9	383	2	A42111	Na+/H+-exchanging
28	111.5	4.9	388	2	B64459	Na+/H+-exchanging
29	111.5	4.9	397	2	S66518	protonase-activat

```

|||||
Db 361 FFESTDIFVIGGIAMSFHGYLSALAMGYTPNVVPSHYSRFAAQLSVCTLMVGLLTGGLW 420
QY 421 PVVIEHFVDKPSIL 434
|||||
Db 421 AVVIEHFVDKPSIL 434

RESULT 2
T23528
hypothetical protein ZK809.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T23528
R:Dobson, R.
submitted to the EMBL Data Library, December 1995
A:Reference number: Z20454
A:Accession: T23528
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-461 <WIL>
A:Cross-references: EMBL:Z68303; PIDN:CAA92642.1; GSPDB:GN00022; CESP:ZK809.4
A:Experimental source: clone ZK809
C:Genetics:
A:Gene: CESP:ZK809.4
A:Map position: 4
A:Introns: 54/3; 110/2; 142/1; 357/3
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK809.4

Query Match 31.2%; Score 716.5; DB 2; Length 461;
Best Local Similarity 35.8%; Pred. No. 4.5e-49;
Matches 163; Conservative 86; Mismatches 167; Indels 39; Gaps 9;
QY 7 SNTYAVEQEA-FPRDKYNIYVWLIVGVGLLPWNMFITIAPEYVYVYWFKPD----- 59
Db 11 NKTAKVEEPEPEKDKGNLVFIILHIGTGLMPWNMLITISDYFESYKMLANSTIDMD 70
QY 60 -----GVETWYSKEFGMGLTIGSOLPNASINVFNLIIAGPLTYRVFAPVCFNIVNLTI 114
Db 71 TGVVVTGYPTVYSSNFQSFOTIASQVPLNLLNLLNIFIVKGLASRI--TVGLSIVAVCV 128
QY 115 ILILVIVLEPTEDSMWFFWTLMGATSNFNSGLYENSIVYGGDPHPHYIGALLIGNN 174
Db 129 ITTMMFIVTSTWLTGFTTITIIIVLNGANGVYQNSIFGLASELPFKYTNNAVIIGNN 188
QY 175 ICGLLIIVWKI---GVTFYFNDEPKLVAIVVFGISLVILLVCAIALFFITKQDFYHVHHQ 231
Db 189 LCGTFVLLSMSTKAVTRNILDR-----SFAYFSIALTLVFCFISFHLKKQRFQYIST 244
QY 232 KGMIREK---AETDRPSPSILMTTTCYQOLFNVWFCFAVTLITPP-VMMTVTTTRGDS 287
Db 245 RAERQRKNKDEAVDSEKGVANYIATPKFAFPQLINVLVFFVTLSTIFPGVMYVKDEKKG 304
QY 288 G-----FLNKIMSENDEI-----YTLTSLFVLNFAAGSIVASKIHPPTRYLK 333
Db 305 GVDYDPLPROFFTNLSLHINREFSEKFMVDTTFLQNVFVAFICGSIVAGKQWPAKPLW 364
QY 334 FAITLRALFIPFFFCNRYVOTRAYPVFFESTDIFVIGGIAMSFHGYLSALAMGYTPNV 393
Db 365 IPVYLRLLYIPFFIFCNLPETSLPVPFFESTWLVIIAASMSFGSGYFSGLAMWYTSKT 424
QY 394 VPSHYSRFAAQLSVCTLMVGLLTGGLMPVVIHFV 428
Db 425 VDPKSAQVAGMMAGFFLISGVSLIFTMTVMKV 459

RESULT 3
T23528
hypothetical protein K09A9.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T23528

```

```

R:Swirnburne, J.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19753
A:Accession: T23528
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-450 <WIL>
A:Cross-references: EMBL:Z79601; PIDN:CAB01882.1; GSPDB:GN00028; CESP:K09A9.3
A:Experimental source: clone K09A9
C:Genetics:
A:Gene: CESP:K09A9.3
A:Map position: X
A:Introns: 59/3; 115/2; 147/1; 318/1; 346/3
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK809.4

Query Match 30.8%; Score 706.5; DB 2; Length 450;
Best Local Similarity 36.0%; Pred. No. 2.7e-48;
Matches 157; Conservative 88; Mismatches 164; Indels 27; Gaps 8;
QY 12 VEQEA-FPRDKYNIYVWLIVGVGLLPWNMFITIAPEYVYVYWF-----KPD 59
Db 21 VEETPEPEKDKGNLVFIILHIGTGLMPWNMLITISDYFESYKMLANSTIDMDTKV 80
QY 60 GVETWYSKEFGMGLTIGSOLPNASINVFNLIIAGPLTYRVFAPVCFNIVNLTIILV 119
Db 81 GDPTVYSSNFQSFOTIASQVPLNLLNLLNIFIVKGLAGRI--TVGLSIVAVCVITMI 138
QY 120 IVLEPTEDSMWFFWTLMGATSNFNSGLYENSIVYGGDPHPHYIGALLIGNICGLL 179
Db 139 FIVETSTWLTGFTTITIIIVLNGANGVYQNSIFGLASELPFKYTNNAVIIGNLCGTF 198
QY 180 ITVVKI---GVTFYFNDEPKLVAIVVFGISLVILLVCAIALFFITKQDFYHVHHQKMEI 236
Db 199 VTLLSMSTKAVTRNILDR-----SFAYFSIALTLVFCFISFHLKKQRFQYSTRERQ 254
QY 237 R---EKAETDRPSPSILMTTTCYQOLFNVWFCFAVTLITPP-VMMTVTTTRGDSGFLNK 292
Db 255 RAKNEEAADNEGRMANYIATPKFAFPQLINVLVFFVTLSTIFPGVMYVKDEKKGTYDF 314
QY 293 IMSENDEIYTLTSLFVLNFAAGSIVASKIHPPTRYLKFAILRALFIPFFFCNRY 352
Db 315 PLPQN---YFMDVTTTFLQNVFVAFICGSIVAGKQWPAKPLWIPVYLRLLYIPFFIFCN 372
QY 353 VQTRAVPVFFESTDIFVIGGIAMSFHGYLSALAMGYTPNVVPSHYSRFAAQLSVCTLMV 412
Db 373 PETRLPVFFESTWLVIIAASMSFGSGYFSGLSMYSKTVDPSKAQVAGMMAGPFLIS 432
QY 413 GLITGGLMPVVIHFV 428
Db 433 GIVSGLIFTMVIRFV 448

RESULT 4
T23247
hypothetical protein K02E11.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T23247
R:McMurray, A.
submitted to the EMBL Data Library, July 1996
A:Reference number: Z19715
A:Accession: T23247
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-800 <WIL>
A:Cross-references: EMBL:Z77665; PIDN:CAB01223.1; GSPDB:GN00023; CESP:K02E11.1
A:Experimental source: clone K02E11
C:Genetics:
A:Gene: CESP:K02E11.1
A:Map position: 5
A:Introns: 63/3; 80/3; 120/3; 137/1; 178/1; 280/3; 333/3; 418/3; 464/2; 496/1; 521/2;

```



```

Query Match          25.0%   Score 574.5;   DB 2;   Length 800;
Best Local Similarity 32.7%;   Pred. No. 1.3e-37;
Matches 134;   Conservative 88;   Mismatches 165;   Indels 23;   Gaps 10;

QY 19 RDKNYIYWLIVLGVGVLPPWNNFTIATPEYYNYNFK-PDGVETWYSKFMSGLTIGS 77
Db 388 KDLGNYVFFIFMFGFGLLPWNNFLNISFDYITMFKLRSDAGNATWYSSNQMSMTISA 447
QY 78 QLPNASINVENFLITIIAGPLIYRFAVCFENIVNLITILIVILEPTSSMFFFWTL 137
Db 448 QIPSLVESVINIFTAVKGLTRGM--KICLIVQLMVIVTVFVIDTSTWIAFESMTL 505
QY 138 GMATISINFSNGLSYNYGVGDSPHYTYIGALLIGNNICGLLITIVVTKGVTYFVLNDEPKL 197
Db 506 GTIVLNAANGLFQNSMFGLASPEFPKYTNVAVIIGQFCGTAFTVLSM-LTKAASDDVQM 564
QY 198 VAIVYFGISLVILLVCAIALEFFITKQDFYHYHKGMEIREKAETDRSPSILWTT---- 253
Db 565 RASLFFGLSSVAVVVCIFLLNFKLAFY----KKFGILRTSSOSDERGIS-SWESVKLA 619
QY 254 FTNCGYGLFNWVFCFAYTLTTFPPVMTVTTRGDSGEFLNKIMSENDEIYTLTSLFLVFNLF 313
Db 620 FEKSKMQFANIFLVFFVTALFPNVMYVDAKKGELHSFVVP-EKTFMDVVVTFELFNLF 678
QY 314 AAIGSIVASIKHNPTRPYLKAFAILRALFIPFFFFCNVRVQ--TRAYPVFFESTDIFVIG 371
Db 679 AFLGSLMANNIRPGNTVVICVAAREWFEMFYFPAANYHPMDFPRAYPVLFESTWLF 738
QY 372 GIAMFSGHYLSALAMGYTPNVVPSH---YSRFAAQLSVCTLMVGLLTG 417
Db 739 ICIFALTSGYLLSIMMYAPR--SHEDPKIQRMAGMIASFFLIFGIVAG 785

RESULT 5
T01489
Hypothetical protein FL707.13 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 22-Oct-1999
C:Accession: T01489
R:Vysotskaia, V.S.; Schwartz, J.R.; Toriumi, M.; Yu, G.; Kwan, A.; Oji, O.; Liu
rtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Feder
submitted to the EMBL Data Library, June 1998
A:Description: Arabidopsis thaliana chromosome 1 BAC FL707 sequence.
A:Reference number: Z14334
A:Accession: T01489
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-428 <YVS>
A:Cross-references: EMBL:AC003671; NID:q2833627; PID:g3176684; GSPDB:GN00059; A
A:Experimental source: cultivar Columbia
C:Genetics:
A:Gene: ATFL707.13
A:Map position: 1
A:Introns: 185/1

```

[illegible]

Qy 193 DEP---KLVAIVYEGISLVILLVCAIAFLFITTKODFYHHQKGM-----IREKAEE----T 242
Db 201 QDPGLRKSNLYFAVGIVGVVVAVICAFVYNVAHKLPIVKFHEERKNEELIREKSBEKSLT 260

Qy 243 DRPSPLTWTTTCYCOLFNWFCFAVTLTIPVMVTTRGDSGFLNKIMSENDEIYT 302
Db 261 GLAWRTLLDWIVTKVKGSHGFVLLVWTLSIFGYIT-----EDVHS 303

Qy 303 -LRSF-----LVNFLEAAIGSVASKIHWPTRYLKFAILLRALFIPEFFFCNYRVQT 355
Db 304 ELLTDWYPILLIAAYNVFDLVGKCLTAVFMELEDEKIAVGGSIALRLFPYLFWGCLHG-- 360

Qy 356 RAYPVFSTEDIFVIGGIAMFSFGHYLSALAMGYTPNNWFSHYSRFAQLSVCTLMVGLL 415
Db 361 ---EMFURTEPVPVTLCTLGLTNGYLTSVLMLAPKSVPLRHSETAGIVTVMFVVGLA 417

Qy 416 TGGI--WPVVI 424
Db 418 SGSVIAMFWVI 428

RESULT 6
C86156
hypothetical protein AAG10625.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C86156
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719
A:Accession: C86156
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-408 <STO>
A:Cross-references: GB:AE005172; MID:99972375; PIDN:AAG10625.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 10.9%; Score 249.5; DB 2; Length 408;
Best Local Similarity 22.08; Pred. No. 2.5e-12;
Matches 99; Conservative 72; Mismatches 187; Indels 93; Gaps 14;

Qy 12 VEQAFAFRDKYNIYVWLIVLGVGLLPWNFFITAIPEYYVNWFKPDGYETWSKEFMG 71
Db 7 IVDEVTRDAYRAYVYTHFLGASLIPLNALITAVD--YGYLYPKHKVEKTFTYVM- 63

Qy 72 SLTIGSOLPNASINVFNLFLIIAGPLIYRFAPYCFNVLNTIILIVILEPTEDSMWS 131
Db 64 -----SCSVLVVLMTWNTRMRYVRMNLGFS-----MFIAMTSLPID--W 105

Qy 132 F-----FWVTLGATSIENSGLYENSYGVDGPHTTYIGALLIGN----- 174
Db 106 VWKGEKENSYMVMVGSVVLGLADGVGSLIGSACKLPQYMQAIFAFASTASSKPFS 165

Qy 175 -----ICGLLTVVKIGVTFLENDEP---KLVAIVYEGISLVILLVCAIAIFF 219
Db 166 LLCKTVLLSLFLGLIIISLRITAKSLPOTPOGMTSRSAHSYFIVSSTILLCCFISCNV 225

Qy 220 ITKODFYHYH---HOKGMEIREKAETDRPSSIL--WTFTNCYQLFNWFCFAVTLTI 274
Db 226 LHKLPMVOQHLLKFHQ-----PLHSTLTIMWGRKIKNPASGMLIITSVTLSI 272

Qy 275 FPVMVTTRTGDSGFL-NKIMSENDEITYTLTSLFVNLFMAAIGSIVASIKHMPTPRYLK 333

Query Match	10.9%	Score	249.5	DB 2	Length	408			
Best Local Similarity	22.0%	Pred. No.	2.5e-12						
Matches	99	Conservative	72	Mismatches	187	Indels	93	Gaps	14
Qy	12	VEQAFAPDKYNIYVWLIVLFGVLLPWNNFIITAIPEYVYNYWFKPGCVETWYSKEPMG	71						
	:		:		:		:		:
Db	7	IVDEVETRDYRAYVYIHFLLGAGSLIPWNNALITAVD--YFGYLYPDKHVKEKTFTVAYM-	63						
Qy	72	SLTTCGSQLPNASINVENFLIIAGPLIYRVAPCFNIVNLTITLILVIVLEPTEDSMW	131						
	:		:		:		:		:
Db	64	-----SCSVLVLMWMTWNTMSYVRNMLGFS-----MFIAMMTSPILID--W	105						
Qy	132	F-----FWTLGMATSNFNSGLYENSIVYGVGGDFPHTYIGALLIGNN-----	174						
	:		:		:		:		:
Db	106	VWKGEKGNVSVMWGVSVLGLADGVVGGSLIGSAGKLPQWMAIFAGTASGKPSF	165						
Qy	175	-----ICGLLTIVYVKGIVTYFNLDEP---KLVAIVYGISVILLVCAIALFF	219						
	:		:		:		:		:
Db	166	LLCKTLVLLSLFLGIIISLRATKASLPQPMQRTSAHSYFIVSVTILLCFCISCNV	225						
Qy	220	ITKODFYVHV---HOKGMREIKAEKTRDPSIL--WTFNTCYGQLENVWFCFAVTLTI	274						
	:		:		:		:		:
Db	226	LHLKPVMOQHLKFHQ-----PLHSLTIWVGRIKWPASGMLIIYSVTLSI	272						
Qy	275	FPVMMVTTRGDSGFL-NKIMSENDEIYTLTSLVFNLFALGASIVASKIHWPTPYLK	333						
	:		:		:		:		:

RESULT 7

A: Title: A

A: Residues

C; Comment:

A; start co

F;107/Binding site: phosphate (Thr) (covalent) (by casein kinase II) #status predicted

Matches	71	Conservative	52	Mismatches	139	Indels	60	Gaps	8
Matches	71	Conservative	52	Mismatches	139	Indels	60	Gaps	8

Db 7 INFSAVLOGSLFGOLGTMPTSTYTLFLSGOGLAGIFAAALAMLLSMASGV-----DAET 60

61 SAIGYFTTPYVGILMSTVCYISLPHIKFARYYLANKSSOAOAOELETAKELIOSDENGIP 120

121 SSPOKVAIT.I.DI.DI.EKEPESEFPDPOKPGKPSVFTVEOKTWT.TAL.CI.VI.VFTVTI.SVEPA 180

191 TTAMWTSCMS-----DCKWJSGFENDTCCET I ENTMDWJ CBER TSEVER WPDENCPT I D I I 23A

225 WOI BEI ENVI ENVI CUIVDOBER-I OIT EDOOAYETEMI I ENVCNCVI VCI EMCI ADOOVI 202

DATE	TIME	LOCATION	WIND DIRECTION	WIND SPEED	WAVE HEIGHT	SEA STATE	TEMPERATURE	MOON	PHASE	REMARKS
2002	10:00	1000	100	10	10	10	10	10	10	10
2002	11:00	1100	110	11	11	11	11	11	11	11
2002	12:00	1200	120	12	12	12	12	12	12	12
2002	13:00	1300	130	13	13	13	13	13	13	13
2002	14:00	1400	140	14	14	14	14	14	14	14
2002	15:00	1500	150	15	15	15	15	15	15	15
2002	16:00	1600	160	16	16	16	16	16	16	16
2002	17:00	1700	170	17	17	17	17	17	17	17
2002	18:00	1800	180	18	18	18	18	18	18	18
2002	19:00	1900	190	19	19	19	19	19	19	19
2002	20:00	2000	200	20	20	20	20	20	20	20
2002	21:00	2100	210	21	21	21	21	21	21	21
2002	22:00	2200	220	22	22	22	22	22	22	22
2002	23:00	2300	230	23	23	23	23	23	23	23
2002	24:00	2400	240	24	24	24	24	24	24	24

C, species: mus musculus (house mouse)
C: Date: 03-Sep-1995 #sequence revision 27-Oct-1995 #text change 05-Nov-1999

A;Title: A mammalian delayed-early response gene encodes HNP36, a novel, conserved nuclear

QY 81 NASINVENLFLIAGPL-----IYRVPAPCVFNIVNLTII-----LILVIVLEPT----- 125
Db 62 TELQANFEAYLVGYSIACVLGSLNVPATK--SLSNSTRMIWGHILVVVFIPIALTF 119
QY 126 ---EDSMWFFWTLGMAITSINF-SNGLYENSIVYGVGGDFPHYTGALLIGNICGLLIT 181
Db 120 WNFDDDDQFFFNLSMILSIACFGSLGMLAGVGLSLAFPSQYTOAVMVGQSFAGVLA 179
QY 182 VVKI---GVT---YFUNDPEKVAIVYFGISVILLVCAIALFFIT-----KODF 225
Db 180 LMSILCOAVTSDVILNGQ-----MYFGSLIMCISLTIATYYLTLTPPMITDDGSGEL 233
QY 226 YHYHOKGMEIR-----EKAETDRPSPSILMTFTTNCYGOLENVWFCAVTL 272
Db 234 IENEVEVIEAQAHHFPTDSNSQTEEHQLPK---WTMYTDIIRK-----SAIDL 282
QY 273 TIFPVMIVTTRGDSGFLNKMSEN-----DEIYTLLSFLVFNFAAGISIVASKIHP 327
Db 283 TTISVVLIVTLAAYPGLTSLVHSTSRNHTWNSYFSAVASFLLYNVGDLIGRSSANSRL- 341
QY 328 TPYLKFAIILRALFIPFFFCN-----YRVQTRAYVPVFESTD-----IFVIG 371
Db 342 SPKYLLIISFLRFALIPMIAMCNVHILMHLFMTVLEFLSIDSIDKDSRELAGSIIALF 401
QY 372 GIAMS-----FSHGYSALAMGYT 390
Db 402 GVTAIFISFFCIIHFSCISIIYVT 425
RESULT 10
FUN26 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YAL022c
C:Species: Saccharomyces cerevisiae
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 29-Oct-1999
R:Accession: S36712
R:Ouillet, F.; Clark, M.W.; Keng, T.; Storms, R.K.; Zhong, W.; Zeng, B.; Fortin, N.; D
submitted to the EMBL Data Library, January 1993
A:Description: Sequencing of Chromosome I from Saccharomyces cerevisiae: analysis of a 3
A:Reference number: S36711
A:Accession: S36712
A:Molecule type: DNA
A:Residues: 1-517 <OUE>
A:Cross-references: EMBL:L05146; NID:g171851; PIDN:AAC04935.1; PID:g171853; MIPS:YAL022c
C:Genetics:
A:Gene: SGD-FUN26
A:Cross-references: SGD:S0000020; MIPS:YAL022c
A:Map position: 1L
C:Keywords: transmembrane protein
F:79-95/Domain: transmembrane #status predicted <TM>
F:155-171/Domain: transmembrane #status predicted <TM2>
F:177-193/Domain: transmembrane #status predicted <TM3>
F:218-234/Domain: transmembrane #status predicted <TM4>
F:245-261/Domain: transmembrane #status predicted <TM5>
F:344-360/Domain: transmembrane #status predicted <TM6>
F:412-428/Domain: transmembrane #status predicted <TM7>
F:445-461/Domain: transmembrane #status predicted <TM8>
F:500-516/Domain: transmembrane #status predicted <TM9>
Query Match 8.0%; Score 182.5; DB 2; Length 517;
Best Local Similarity 19.9%; Pred. No. 6.3e-07;
Matches 79; Conservative 69; Mismatches 131; Indels 117; Gaps 20;
QY 23 NIVYWLVLVGVGLLPNMFTITAPYYVNVWFKPDGVETWYSKEFMGSL 79
Db 75 NLSYITFTFAIGLLPWNWNCILS-ASQYFKHDFKDTSI--WAKIFTSSMMSFSTISSML 131
QY 80 PNASINVENLFL-----IIAGPLIYR--VFAPVCENIVNLTIIILVIVLEPTED 127
Db 132 -----FNIIYLAQKQYKYSRVING-LWWEIIVFVMCF-----FTILHFL- 171
QY 128 SMSWFFWTLGMAITSI-NFNSGLYENSIVYGVGGDFPHYTGALLIGNICGLLITVVKIG 186

Db 172 -PKWFNEFMIMLVVVISSMGTAQTQIMAIANVFGESEYSQGVMGVQAVAGVLPVSLVLF 230
QY 187 VTYFINDPEKLVAA--IVYFGISLVILLVCAIALFFITK----- 222
Db 231 LAFIENSSVSTTGGILLVFFFTTLLVVVIC-VVMFVSYSKISRKVRKNWNNVVDGHITDVLIG 289
QY 223 -----QDFVHYH-----HOKGMEIREKAEATDRPSPSILMTTFTNC 257
Db 290 SLRSEEEIRIVGRIDQDEDEHRTNGTRDDNDGGEELQKVPE-----VLFAKL 341
QY 258 YGQLENVWFCFAVTLITIPVMMVITTRGDSGFLNKMSENDEIYTLTSLVFNLFPAAG 317
Db 342 KYLVLSIETFEVVTL-VFVPFASAT-----YVTGLPSNAQYIPLI--FTLWNLGDLYG 392
QY 318 SIVASKIHP-----TPYLKFAIILRALFIPFF 346
Db 393 RVIAD---WPMFRDQKFTPRKFTIYSLLRVAAIPLF 425
RESULT 11
T22164
hypochemical protein F44D12.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22164
R:Coles, L.
submitted to the EMBL Data Library, December 1995
A:Reference number: Z19525
A:Accession: T22164
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-143 <WIL>
A:Cross-references: EMBL:Z68298; PIDN:CAA92605.1; GSPDB:GN00022; CESP:F44D12.9
A:Experimental source: clone F44D12
C:Genetics:
A:Gene: CESP:F44D12.9
A:Map position: 4
A:Introns: 19/1; 37/2
Query Match 7.8%; Score 180; DB 2; Length 143;
Best Local Similarity 40.0%; Pred. No. 2.4e-07;
Matches 32; Conservative 18; Mismatches 26; Indels 4; Gaps 1;
QY 14 QZAFPDPKYNIVYWLVLVGVGLLPNMFTITAPYYVNVWFKPDGVETWYSKEFMGSL 73
Db 67 KESSPTDRWLYVIIFTMTHGMGLMSWNMFITIAPOYHYHWFN-----NTNYODSFMSII 122
QY 74 TIGSQLPNASINVENLFLII 93
Db 123 GVTSQLPNVNGIMLTIVVM 142
RESULT 12
DB5064
hypochemical protein AT4G05120 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse ear cross)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: DB5064
R:anonymouse, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488
A:Accession: DB5064
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-418 <STO>
A:Cross-references: GB:NC_001268; NID:g7267271; PIDN:CAB81054.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4G05120
A:Map position: 4

[illegible][illegible]

Qy	20	DKN-----IVWLVLGCGVLLPWNFEITTAPEYVNYWFKDGVETWYSKPMGSL	73
Db	12	EKHPSRLVITVQLVA-----NVFII-----TLATKE-----	39
Qy	74	TIGSOLPNASINVFNLFLIIAGIPLVYFAPVCENIVLT-----IILILVILE	123
Db	40	---AKLTRLRNIFGYSLYTAGTF-----CLIIIDLASHGSGSWAVYVLLCLIALV-	87
Qy	124	PTEDSMSEFFVWTLGMATSI NFSGYLSVYGVGGDFPHYTI GALLIGNCTGLLITVY	183
Db	88	-----FGLADA--FVQGVAMVDLSFMSPDF---IQAFMAGLIGTAGALTSVL	128
Qy	184	KIGVTYFLNDEP---KLVAIVYEGISILVILLVCAI--ALFFITKQDFYHHYHOKME--	235
Db	129	RLITKAIFDPSOGLRGKALLGIGATLIELACVFLYLVFAKPLVYKIRAKAGKEGAK	188
Qy	236	-----IREKAEITDRPSILMTTFTNCYQGL-----FNWCFCEAVTLTIPFV	277
Db	189	TVSADLAAGLQEQAEQVHQWDESKIQKLTk---KQLLRNIDLIGLSLIYVVVLTSlFP-	245
Qy	278	MMTVTRGDSGFLNKIMSEN---DEIYTLTSLFVLNLPFAAGSIVAS--KIHWPTPYVL	332
Db	246	-----GELYENTGEHRLGDWAPVLVA--MYNGWDAISRFITPSIKPLAMESRKWI	293
Qy	333	KFAIILRALFIPFPFFFCNYVQTRAYVPVPESTDIFVIGGIMAFSGHVSALAMGYTEN	392
Db	294	TVCVVARLLLVAPYF-TAKYAOQWMLFITS-----FGLSNGHLYTCIFSTAPK	343
Qy	393	VVPSHYSRFAAQUSVCTLMV-----GLLTGGLWPVPIEHF	427
Db	344	GYNGPEANALGNL-MCVFELLGIGIFAGVCLGWLLIGNDSF	382

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 16:56:28 ; Search time 14.4 seconds
(without alignments)
678.224 Million cell updates/sec

Title: US-08-816-01lf-63
Perfect score: 2294
Sequence: 1 MVLINRNTYAVDSERPRD.....LTGGLWPVVEHVDKPSIL 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PT05_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	343.5	15.0	456	US-09-058-389A-4	Sequence 4, Appl
2	326.5	14.2	456	US-09-058-389A-2	Sequence 2, Appl
3	214	9.3	326	US-09-058-389A-3	Sequence 3, Appl
4	181	7.9	247	US-09-058-389A-8	Sequence 8, Appl
5	111.5	4.9	398	US-08-097-938-4	Sequence 4, Appl
6	111.5	4.9	398	US-08-476-000-4	Sequence 4, Appl
7	111.5	4.9	398	US-08-472-840-4	Sequence 4, Appl
8	111.5	4.9	398	US-08-476-976-4	Sequence 4, Appl
9	111.5	4.9	398	US-08-474-410-4	Sequence 4, Appl
10	111.5	4.9	398	US-08-486-673B-4	Sequence 4, Appl
11	109.5	4.8	397	US-08-486-673B-63	Sequence 63, Appl
12	108.5	4.7	398	US-08-097-938-6	Sequence 6, Appl
13	108.5	4.7	398	US-08-476-000-6	Sequence 6, Appl
14	108.5	4.7	398	US-08-472-840-6	Sequence 6, Appl
15	108.5	4.7	398	US-08-476-976-6	Sequence 6, Appl
16	108.5	4.7	398	US-08-474-410-6	Sequence 6, Appl
17	108.5	4.7	398	US-08-486-673B-6	Sequence 6, Appl
18	106.5	4.6	395	US-08-097-938-2	Sequence 2, Appl
19	106.5	4.6	395	US-08-097-938-5	Sequence 5, Appl
20	106.5	4.6	395	US-08-476-000-2	Sequence 2, Appl
21	106.5	4.6	395	US-08-476-000-5	Sequence 5, Appl
22	106.5	4.6	395	US-08-472-840-2	Sequence 2, Appl
23	106.5	4.6	395	US-08-472-840-5	Sequence 5, Appl
24	106.5	4.6	395	US-08-476-976-2	Sequence 2, Appl
25	106.5	4.6	395	US-08-476-976-5	Sequence 5, Appl
26	106.5	4.6	395	US-08-474-410-2	Sequence 2, Appl
27	106.5	4.6	395	US-08-474-410-5	Sequence 5, Appl

28	106.5	4.6	395	4	US-08-486-673B-2	Sequence 2, Appl
29	106.5	4.6	395	4	US-08-486-673B-5	Sequence 5, Appl
30	106.5	4.6	399	1	US-08-476-000-61	Sequence 61, Appl
31	106.5	4.6	399	1	US-08-472-840-61	Sequence 61, Appl
32	106.5	4.6	399	2	US-08-476-976-61	Sequence 61, Appl
33	106.5	4.6	399	3	US-08-474-410-61	Sequence 61, Appl
34	106.5	4.6	399	4	US-08-486-673B-61	Sequence 61, Appl
35	105.5	4.6	397	1	US-08-476-000-63	Sequence 63, Appl
36	105.5	4.6	397	1	US-08-472-840-63	Sequence 63, Appl
37	105.5	4.6	397	2	US-08-476-976-63	Sequence 63, Appl
38	105.5	4.6	397	3	US-08-474-410-63	Sequence 63, Appl
39	101	4.4	420	1	US-08-466-103A-2	Sequence 2, Appl
40	96	4.2	473	1	US-08-597-236-13	Sequence 13, Appl
41	96	4.2	473	1	US-08-746-682A-13	Sequence 13, Appl
42	95.5	4.2	336	3	US-08-749-816-2	Sequence 2, Appl
43	94.5	4.1	436	6	5432081-10	Patent No. 5432081
44	94	4.1	418	4	US-09-030-267-5	Sequence 5, Appl
45	93.5	4.1	569	2	US-08-750-723A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-058-389A-4
; Sequence 4, Application US/09058389A
; Patent No. 6130065
; GENERAL INFORMATION:
; APPLICANT: Belt, Judith A.
; APPLICANT: Crawford, Charles R.
; APPLICANT: Patel, Divyen
; TITLE OF INVENTION: A NITROBENZYLMECAPTOPURINERIBOSIDE
; TITLE OF INVENTION: (NBMPR)-INSENSITIVE EQUILIBRATIVE, NUCLEOSIDE TRANSPORT
; TITLE OF INVENTION: PROTEIN, NUCLEIC ACIDS ENCODING THE SAME AND METHODS OF
; TITLE OF INVENTION: USE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/058,389A
; FILING DATE: April 9, 1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-013N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: hENT1
; HYPOTHETICAL: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-09-058-389A-4


```

; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/058,389A
; FILING DATE: April 9, 1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-013N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 326 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: hHNP36
; HYPOTHETICAL: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-09-058-389A-3

Query Match 9.3%; Score 214; DB 4; Length 326;
Best Local Similarity 22.0%; Pred. NO. 1.8e-13;
Matches 71; Conservative 52; Mismatches 139; Indels 60; Gaps 8;

QY 143 INFNSGLYENSVYGGDPHYIGALLIGNNICGLITVVKI-----GVTYFLNDEPKL 197
DB 7 INFSAVLQSGFLGQGTWMPSTYTLFSGQGLAGIFAALMLLSMASGV-----DAET 60

QY 198 VAIVYF---GISLVLLVCAIALFFITKODFY---HYHHQKMEIREKA-----240
DB 61 SALGYEITPYVGIILSVYCLSLPLKFKARYLANKSSQAQAQAELETRAEQLSDENGIP 120

QY 241 -----ETDRSPSILWTTFTNCYQLNWFECFPAVTLTIPV 277
DB 121 SSPQKVALTLDLDEKEPESEDEPKGPKPSVFTVFQKIWLTALCLVLVFTVTLSEFPA 180

QY 278 MMTVTTRGDSGFLNKIMSENDEIYLLTSLFVNLFAAIGSIVASKIHP--TPRYLKFA 335
DB 181 ITAMVTSST-----PGKWSOFFNPICCFLEFNIMDLGRSLTSYFLWPDEDSRLPL 234

QY 336 IILRALFIPFFFCNVYOTRAYPVFFESTDIFVIGGTAMSFSGYLSALAMGYTPNVVP 395
DB 235 VCLRELFVPLEMLCHVPQSR-LPILFPQDAVFTFMLLFAVNGVLSVLSLTCAPQVYL 293

QY 396 SHYSRFAQLSVCTLMVGLTG 417
DB 294 PHEREVAGALMTFTFLALGLSCG 315

RESULT 4
US-09-058-389A-8
; Sequence 8, Application US/09058389A
; Patent No. 6130065
; GENERAL INFORMATION:
; APPLICANT: Belt, Judith A.
; APPLICANT: Crawford, Charles R.
; APPLICANT: Patel, Divyen
; TITLE OF INVENTION: A NITROBENZYL MERCAPTOPURINERIBOSIDE
; TITLE OF INVENTION: (NBMPR)-INSENSITIVE, EQUILIBRATIVE, NUCLEOSIDE TRANSPORT

; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/058,389A
; FILING DATE: April 9, 1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-013N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 326 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: hHNP36
; HYPOTHETICAL: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-09-058-389A-3

Query Match 7.9%; Score 181; DB 4; Length 247;
Best Local Similarity 23.6%; Pred. No. 2.5e-10;
Matches 59; Conservative 42; Mismatches 107; Indels 42; Gaps 7;

QY 18 PRDKYNIVVWLVLVGFVLLPWNMFITIAPEY-----VNTWFKPDGVE 62
DB 7 PRDYLHVGISFELLGLGTLLEPWFITAIPTFYQARLAGAGNSTARILSTNHTGPDFAF- 65

QY 63 TWYSEFMGSLTIGSQLPNASINVENFLIIAGPLIYRVFAPVCNIVNLTILILVIVL 122
DB 66 -----NFNNWVTLSSQLPLLETLNLSFLYQCVETVRILG----SLLAILLFALTAAL 116

QY 123 EPTEDSMWFFWVTLGMATSFNSNGLYENSVYGGDPHYIGALLIGNNICGLITV 182
DB 117 VKVDMSPGPFSTWASVCFINSFSAVLQSLFGQLGTPWSTYTLFSGQGLAGIFAAL 176

QY 183 VKI-----GVTYFLNDEPKLVAIVYF---GISLVLLVCAIALFFITKODFY---HYHHO 231
DB 177 AMLLSMASGV-----DAETSALGYFITPCVGIILSVYCLSLPLKFKARYLANKSSQA 230

QY 232 KCMETREKAE 241
DB 231 QAQELETKAE 240

RESULT 5
US-08-097-938-4
; Sequence 4, Application US/08097938
; Patent No. 5629174
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR AND ITS
; TITLE OF INVENTION: AGONISTS AND ANTAGONISTS
; NUMBER OF SEQUENCES: 59
```



```

Db      329 VALCLSTLNSCIDPFVYFVSHDFRDHAKNALLCRSVRTVKQMOV 373
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT      8
US-08-476-976-4
; Sequence 4, Application US/08476976
; Patent No. 5874400
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,976
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/390,301
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 2803-0006.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PS-08-476-976-4

```

```

Query Match      4.9%; Score 111.5; DB 2; Length 398;
Best Local Similarity 18.5%; Pred. No. 0.0041;
Matches 75; Conservative 54; Mismatches 131; Indels 145; Gaps 18;

y   61 VETWYS-KEFMGSLTIGSOLPNASINVNFLIIAGPLIYRVFAPVCNIVNLTILLV 119
    ||| : | : | : | : | : | : | : | : | : | : | : | : | : |
o   56 VETFVSDFESAVLTGK-----LTTVFLLPVTIV-----F 87
    ||| : | : | : | : | : | : | : | : | : | : | : | : | : |
y   120 IVLEPTEDSMSFF-----NWTLGMATSI NFSNGLYENS 153
    ||| : | : | : | : | : | : | : | : | : | : | : | : | : |
o   88 VYGLPSNGMALWFELRTKKHPAIVYMANIALADLLSVIFPLKIATHIGNNNIYGEA 147
    ||| : | : | : | : | : | : | : | : | : | : | : | : | : |
y   154 VYGVGGDPHTYIGALLIGNNICGLITVWKIGTVFYLNDEP----KLVAIVYFGISLVI 209
    ||| : | : | : | : | : | : | : | : | : | : | : | : | : |
o   148 LCNV-----LIG-FFYGNMYCSILPMTCLSVQRYWYIVNPMGHSRKKANIAIGISLA I 199
    ||| : | : | : | : | : | : | : | : | : | : | : | : | : |
y   210 -LLVCAIAL-FEITKQDFYHYHHQKMEIREKAETDRPSILMTFTFNCY----- 258
    ||| : | : | : | : | : | : | : | : | : | : | : | : | : |
o   200 WLLILLVTPLYVVQRQTF-----IPAL--NITTCHDVLPEQLLV 237
    ||| : | : | : | : | : | : | : | : | : | : | : | : | : |
y   259 GOLFNWFCFAVTLTIFFVMVTVTTTRGDSGFL-----NKIMSENDE-----IYT 302
    ||| : | : | : | : | : | : | : | : | : | : | : | : | : |
o   238 GDMENYFLSLAIQVLFEPAPLFA-----SAYVLMIRLSSAMDSENSKKRAIKFLIVT 292
    ||| : | : | : | : | : | : | : | : | : | : | : | : | : |

```


RESULT 11
US-08-486-673B-63
; Sequence 63, Application US/08486673B
; Patent No. 6297026
; GENERAL INFORMATION:
; APPLICANT: Sundelin, Johan
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: Nucleic Acids Encoding the C140 Receptor
; FILE REFERENCE: 44481-5006-08-US
; CURRENT APPLICATION NUMBER: US/08/486,673B
; CURRENT FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/097,938
; PRIOR FILING DATE: 1993-07-26
; PRIOR APPLICATION NUMBER: PCT/US94/08536
; PRIOR FILING DATE: 1994-07-26
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-486-673B-63

Query Match 4.8%; Score 109.5; DB 4; Length 397;
Best Local Similarity 19.0%; Pred. No. 0.0065;
Matches 76; Conservative 57; Mismatches 131; Indels 137; Gaps 19;
QY 61 VETWYS-REFMGLSLIGSOLPNASINVENFLIAGPLIYR-VFA----- 103
DB 55 VETVFSVDEFSASVLG-----KLTTFVL-----PIVYTIIVFVGLPSNGMALWFLF 102
QY 104 -----PVCFNIVNTIILVILEPTEDSMWFFWTLGWTATSNFNGLYENSVYGV 157
DB 103 RTKKKHPAVIYMANLADLLSVI-----WFLPKIAYHHGNWNYGALCNV 150
QY 158 GGDFFHTYIGALLNIGCGLLITVVKIGVYFLNDEP-----KLVAIVYFGISLVILVLC 213
DB 151 -----LIG-FYGNMYSILEMTCLSVQRYVWVNPVGMHSHKRAKNAIGSLAIWLLT 202
QY 214 ATAL--FFITKDFHYHQQKMEIREKAEIDRPSILWTTFTNCTCY-----GQLF 262
DB 203 LLVITPLVYVVKQITF-----IPAL---NITTCDDVLPQQLVGDGF 240
QY 263 NVWFCFAVLTIFPVYMTVTRGDSGFL-----NKIMSENDE-----IVTLTTS 306
DB 241 NYFLSLAIGVFLFPAFLTA-----SAYVLMIRLSSAMDENSEKKRRAIKLIIVTVLCM 295
QY 307 FLVFNLFRAAGSIVASKIHWPTPRYLKFAILRALFIPFFFCNRYVQTRAYVPFFESTD 366
DB 296 YLI--CF-----TPSNL-----LLVVHYFLIKSQGSHVYALVIALC 331
QY 367 IFVIGGIAMSFHGYLSALAMGYTPNVVPSHYSRFAAQLSV 407
DB 332 LSTLNSCIDPFYVYFVSHDFRDHAKNALLCRSVRTVKQMV 372

RESULT 12
US-08-097-938-6
; Sequence 6, Application US/08097938
; Patent No. 5629174
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR AND ITS
; AGONISTS AND ANTAGONISTS
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20006-1812

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/097,938
; FILING DATE: 26-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 22803-20006.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-097-938-6

Query Match 4.7%; Score 108.5; DB 1; Length 398;
Best Local Similarity 19.2%; Pred. No. 0.0083;
Matches 75; Conservative 59; Mismatches 141; Indels 115; Gaps 18;
QY 61 VETWYS-REFMGLSLIGSQ-----LPNASINVENFLIAGPLIYRVFA-----PVCFN 108
DB 56 VEIVFSVDEFSASVLTGKLTTFVFLPIVYIIVFVGLPSNGMALW-VFLFRTKKKHPAVIY 114
QY 109 IVNLTIIILVILEPTEDSMWFFWTLGWTATSNFNGLYENSVYGVGGDFPHYIGA 168
DB 115 MANLADLLSVI-----WFLPKIAYHHGNWNYGALCNV-----LIG- 154
QY 169 LLIGNIGCGLLITVVKIGVYFLNDEP---KLVAIVYFGISLVI-LLVCATAL-FFITK 222
DB 155 FFGNMYCSILEMTCLSVQRYVWVNPVGMHSHKRAKNAIGSLAIWLLIIVTPLYVVK 214
QY 223 QDFYHYHQQKMEIREKAEIDRPSILWTTFTNCTCY-----GQLFNWFCFAVLT 273
DB 215 QTFI-----IPAL---NITTCDDVLPQQLVGDGFENYFLSLAIGVF 252
QY 274 IFPVYMTVTRGDSGFL-----NKIMSENDE-----IYTLTSLFVFNLFALIG 317
DB 253 LEPALFTA-----SAYVLMIRLSSAMDENSEKKRRAIKLIIVTVLAMYLI----- 299
QY 318 SIIVASKIHWPTPRYLKFAILRALFIPFFFCNRYVQTRAYVPFFESTDIFVIGGIAMSF 377
DB 300 CFIPSNL-----LLVVHYFLIKSQGSHVYALVIALCLSTLNSCIDPF 343
QY 378 SHGYLSALAMGYTPNVVPSHYSRFAAQLSV 407
DB 344 VYFVSHDFRDHAKNALLCRSVRTVKQMV 373

RESULT 13
US-08-476-000-6
; Sequence 6, Application US/08476000
; Patent No. 5716789
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington

Sequence 6, Application US/08476976
Patent No. 5874400
GENERAL INFORMATION:
APPLICANT: SUNDELIN, JOHAN
APPLICANT: SCARBOROUGH, ROBERT M.
TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476,976
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/390,301
FILING DATE: 25-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: ADLER, REID G.
REGISTRATION NUMBER: 30,988
REFERENCE/DOCKET NUMBER: 2803-0006.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 398 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-476-976-6

Db 344 VYFVSHDFRDHAKNALLCRSVRTVKQMOV 373
Search completed: February 27, 2002, 16:59:53
Job time: 205 sec

Query Match 4.7%; Score 108.5; DB 2; Length 398;
Best Local Similarity 19.2%; Pred. No. 0.0083;
Matches 75; Conservative 59; Mismatches 141; Indels 115; Gaps 18;

QY 61 VETWYS-KEFMGSLTIGSQ-----LPNASINVFNLIIAGPLIYRVFA-----PVCEN 108
Db 56 VEIVSFVDEFSASVLTGKLTTPVLPVYIIVFVGLPSNGMALW-VFLFRTKKKHPAVIY 114
QY 109 IVNLTILIVLIVLEPTEDSMFVWVTIGMATSNFNSGLYENSIVYGVGDPPHYTICA 168
Db 115 MANLADLLSVI-----WFLKIAVTHGNWIIYGEALCNV-----LIG- 154
QY 169 LLIGNNICGLITVVKIGTYFINDEP----KLVAIVYFGISLVI-LLVCAIAL-FFITK 222
Db 155 FFVGNMYCSILEMTCLSVQRYVIVNPMGHSRKKANIAIGISLAIWLLILLVTIPLYVVK 214
QY 223 QDFYHYHOKGMEIRKAEKTDPRSPILWTTFTNCF-----GOLFNVWFCFVTLT 273
Db 215 QTIF-----IPAL-----NITTCVDLPEQVLVGDMMENYELSLAIGVF 252
QY 274 IFVMMVTVTTRGDSGFL-----NKIMSENDE-----IYTLTSTFLVLENFAAIG 317
Db 253 LFPAPLTA-----SAYVLMIRLRSAMDENSEKKRAIKLIVTVLAMYLI----- 299
QY 318 SIVASKIHNPTRYLKFAILRLAFIPFFFCNRYVOTRAYPVFEESTDIEFVIGIAMSF 377
Db 300 CFIPSNL-----LLVVHYFLIKSQCSQSHVYALYIVALCLSTLNSCIDPF 343
QY 378 SHGYLSALAMGYTPNVVPSHYSRFAAQLSV 407

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 16:56:13 ; Search time 25.35 seconds
(without alignments)
1268.158 Million cell updates/sec

Title: US-08-816-011f-63
Perfect score: 2294
Sequence: 1 MVLINRNTYAVEQEAFPRD.....LTGGLWPVVIEHVDKPSIL 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2290	99.8	434	17 AAR92315	CORK potassium cha
2	372.5	16.2	475	21 AAY82286	Rat EN1 receptor
3	343.5	15.0	456	19 AAW69555	Human equilibrativ
4	343.5	15.0	456	21 AAB15520	Human EN1 protein
5	327.5	14.3	475	21 AAY82285	Human EN1 recepto
6	326.5	14.2	456	19 AAW83965	Equilibrative nucl
7	326.5	14.2	456	19 AAW69558	Human equilibrativ
8	326.5	14.2	456	21 AAB15517	Human NEMPR-iENTP
9	326.5	14.2	475	21 AAY93363	Human PRO1380 (UNQ
10	326.5	14.2	475	22 AAB66112	Protein of the inv
11	320.5	14.0	456	19 AAW69557	Rat equilibrative

12	301	13.1	457	19 AAW69556	Rat equilibrative
13	272.5	11.9	373	19 AAW64550	Human fibrosarcoma
14	261.5	11.4	397	22 AAB93550	Human protein sequ
15	214	9.3	326	21 AAB15519	Human HNP36 protei
16	181	7.9	247	19 AAW83966	Alternatively spli
17	181	7.9	247	21 AAB15518	Human NEMPR-iENTP
18	180	7.8	314	20 AAB88377	Polypeptide fragme
19	179.5	7.8	423	22 AAB88377	Human membrane or
20	119.5	5.2	383	22 AAG82466	S. epidermidis ope
21	114.5	5.0	527	21 AAY76126	Human secreted pro
22	113.5	4.9	329	22 AAB92687	Human protein sequ
23	113.5	4.9	360	22 AAB75546	Human secreted pro
24	113.5	4.9	530	22 AAB75598	Human secreted pro
25	113.5	4.9	545	22 AAB94667	Human protein sequ
26	113.5	4.9	579	21 AAY82460	Human SM-11044-bin
27	113.5	4.9	582	21 AAY82459	Human SM-11044-bin
28	111.5	4.9	397	21 AAB35641	Human PAR-2 protei
29	111.5	4.9	398	16 AAR66921	Human C140 recepto
30	111.5	4.9	398	17 AAW01953	Human C140 recepto
31	111.5	4.9	545	21 AAY94910	Human secreted pro
32	111	4.8	439	19 AAW61371	Non-adrenergic SM
33	108.5	4.7	776	21 AAY96167	Schizosaccharomyce
34	108	4.7	354	18 AAW25927	Xenopus melatonin
35	106.5	4.6	392	22 AAE06597	Human protein havi
36	106.5	4.6	399	16 AAR66922	Murine C140 recept
37	106.5	4.6	399	17 AAW01954	Murine C140 recept
38	106.5	4.6	439	22 AAG82657	S. epidermidis ope
39	106	4.6	318	22 AAG71666	Human olfactory re
40	106	4.6	318	22 AAG72405	Human OR-like poly
41	106	4.6	465	22 AAB95721	Human protein sequ
42	105.5	4.6	397	16 AAR66923	Human C140 recept
43	105.5	4.6	397	17 AAW01955	Human C140 recept
44	105.5	4.6	438	22 AAB60107	Human transport pr
45	105	4.6	393	21 AAY81642	Streptococcus pneu

ALIGNMENTS

RESULT 1
AAR92315
ID AAR92315 standard; Protein: 434 AA.

XX AAR92315;

XX AAR92315;

DT 16-OCF-1996 (first entry)

XX CORK potassium channel protein.

DE CORK; potassium channel; nematode; pore-forming domain;

XX transmembrane helix; N-glycosylation site; potassium-agonist;

KW potassium-antagonist; drug screening; nematocide; anthelmintic;

KW cardiac disorder.

XX Caenorhabditis elegans.

OS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX PD 09-MAY-1996.
 XX PF 25-OCT-1995; 95WO-US14364.
 XX PR 31-OCT-1994; 94US-0322312.
 XX PA (AMCY) AMERICAN CYANAMID CO.
 XX PI Pausch MH, Price LA;
 XX DR WPI; 1996-239450/24.
 XX DR N-PSDB; AAT18168.
 XX PT Potassium channel genes from Drosophila melanogaster and
 XX PT Caenorhabditis elegans - useful in assaying substances to determine
 XX PT effects on cell growth, and in inhibiting nematode and insect pests
 XX PS Claim 13; Fig 9; 79pp; English.
 XX CC This potassium channel sequence is encoded by the CORK gene from
 XX CC Caenorhabditis elegans, and has 2 pore-forming domains situated
 XX CC between hydrophobic transmembrane helix domains. The CORK sequence
 XX CC contains structural features resembling pore-forming H5 domains
 XX CC found in potassium channels. 2 putative pore-forming H5 domains
 XX CC contain the G-Y/F-G tripeptide motif required for potassium
 XX CC selectivity. The protein contains 3 asparagine-linked glycosylation
 XX CC sites. The protein may be expressed in a heterologous host cell to
 XX CC assay substances to determine effects on cell growth. Potassium-
 XX CC agonists or potassium-antagonists identified by this method may be
 XX CC used as nematocides, anthelmintics or in therapy of cardiac
 XX CC disorders, etc.
 XX SQ Sequence 434 AA;

Query Match 99.8%; Score 2290; DB 17; Length 434;
 Best Local Similarity 99.8%; Pred. No. 9.3e-228;
 Matches 433; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVLINRNTYAVEGEAFPRDKYNIVYLVILVGFVLLPNNMTITTAPEYVYVWKPDPG 60
 DB 1 MVLINRNTYAVEGEAFPRDKYNIVYLVILVGFVLLPNNMTITTAPEYVYVWKPDPG 60
 QY 61 VETWYSKEFGSLTIGSQLPNASINVFNLFIAGPLIYRVFAPVCFNVLNLTILVI 120
 DB 61 VETWYSKEFGSLTIGSQLPNASINVFNLFIAGPLIYRVFAPVCFNVLNLTILVI 120
 QY 121 VLEPTDSMSWFFWTLGMAISINFSNGLYENSIVYGVGDFPHYTGALLIGNICGLLI 180
 DB 121 VLEPTDSMSWFFWTLGMAISINFSNGLYENSIVYGVGDFPHYTGALLIGNICGLLI 180
 QY 181 TVVKIGVYFLNDEPKLVAIVYGISLVILLVCAIALFFITKQDFYHHQKMEIREKA 240
 DB 181 TVVKIGVYFLNDEPKLVAIVYGISLVILLVCAIALFFITKQDFYHHQKMEIREKA 240
 QY 241 ETRPSPSILWTFITTCYGLFVNWVCFVATLTIFPVMVTVTTRGDSGLNKMSENDEI 300
 DB 241 ETRPSPSILWTFITTCYGLFVNWVCFVATLTIFPVMVTVTTRGDSGLNKMSENDEI 300
 QY 301 YTLTSLPVLNFAAGISIVASIKHWPTRPYLKFAILRALFIPFFFCNRYVOTRAYPV 360
 DB 301 YTLTSLPVLNFAAGISIVASIKHWPTRPYLKFAILRALFIPFFFCNRYVOTRAYPV 360
 QY 361 FFESTDIFVIGGTAHSFHCYLSALANGYTPNVVPSHYSFAQLSVCTLMVGLLTGGLW 420
 DB 361 FFESTDIFVIGGTAHSFHCYLSALANGYTPNVVPSHYSFAQLSVCTLMVGLLTGGLW 420
 QY 421 PVVIEHFVDKPSIL 434
 DB 421 PVVIEHFVDKPSIL 434

RESULT 2
 AAY82286
 ID AAY82286 standard; Protein; 475 AA.
 XX AC AAY82286;
 XX DT 19-JUN-2000 (first entry)
 XX DE Rat ENT1 receptor SEQ ID NO:5.
 XX KW Rat; ENT1; ENT1; immunosuppressant; vasotropic; thrombolytic;
 KW cytostatic; hypotensive; antiinflammatory; analgesic; anticoagulant;
 KW nucleoside transporter; ischaemia; cerebral embolism; malignant tumour;
 KW organ transplant rejection; renitis; pancreatitis; hypertension;
 KW analgesic; blood platelet coagulation inhibitor.
 XX OS Rattus sp.
 XX PN WO200012550-A1.
 XX PD 09-MAR-2000.
 XX PF 26-AUG-1999; 99WO-JP04602.
 XX PR 27-AUG-1998; 98JP-0241248.
 XX PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX PI Miyaji H, Mimura H, Kambe M, Nakagawa S;
 XX DR WPI; 2000-256588/22.
 XX DR N-PSDB; AAA08030.
 XX PT Novel nucleoside transporter polypeptide, useful in drug development
 PT for agents in treating e.g. ischemia, cerebral embolism, rejection in
 PT organ transplant, malignant tumors, renitis, pancreatitis and
 PT hypertension
 XX PS Example 3; Page 81-83; 94pp; Japanese.
 XX CC The present invention describes human and rat ENT1 receptors (ENTR1),
 CC with some amino-acids deleted, substituted or added, but which retain
 CC nucleoside transport activity, and can be used in drug development for
 CC agents in treating e.g. ischaemia, cerebral embolism, rejection in organ
 CC transplant, malignant tumours, renitis, pancreatitis and hypertension.
 CC The ENT1 proteins are nucleoside transporters which have vasotropic,
 CC immunosuppressant, thrombolytic, cytostatic, hypotensive, analgesic,
 CC antiinflammatory and anticoagulant activities. The transporter proteins,
 CC their DNA, oligonucleotides, transformants and antibodies can be used
 CC in drug development for agents in treating ischaemia, cerebral embolism,
 CC rejection in organ transplant, malignant tumours, renitis, pancreatitis
 CC and hypertension, and as analgesics, or blood platelet coagulation
 CC inhibitors, as well as agents to reduce side-effects during
 CC chemotherapy. The present sequence represents the rat ENT1 protein.
 XX SQ Sequence 475 AA;

Query Match 16.2%; Score 372.5; DB 21; Length 475;
 Best Local Similarity 25.6%; Pred. No. 7.5e-30;
 Matches 120; Conservative 79; Mismatches 209; Indels 61; Gaps 13;

QY 7 SNTYAVEQBFAF-----PRDKYNIVYLVILVGFVLLPNNMTITTAPEYV 52
 DB 21 snrheadqallgkpldypagqlrpedrfgngayiffclgigllpwnffvt-akeya 79
 QY 53 NYWFK-----PDGVETWYSKEFGSLTIGSQLPNASINVFNLFIAGPLIYRV 102
 DB 80 ---fkircnsspasgkdpdadilnyfesyavastvpsllflvanflvnrivrvl 136
 QY 103 APVCFNIVNLTILILVILEPTEDSMW---FFWVTLGMAISINFSNGLYENSIVYGVG 159
 DB 137 aslsvalaifvrmavirv----dtsswtrgrffsiamacmailssstsfnsvvgltg 191

therapy -
 Disclosure; Fig 2A-B; 55pp; English.
 The invention relates to the isolation of a gene encoding a NBMPR-iENTP (nitrobenzylmercaptopyrineriboside-insensitive, equilibrative nucleoside transporter protein) from Hela cells. The protein transports nucleosides across the plasma membrane by a facilitated diffusion process and is involved in the proliferative response. The transporter is useful in screening to identify natural nucleoside permeants and/or their inhibitors or analogues, potential therapeutic agents, also for studies on protein structure and mechanism. Cells that express the iENTP as the only transporter protein are used for drug screening (especially to identify antitumor and antiviral nucleoside analogues), in chemotherapy of cancer and for selective expression of heterologous genes for gene therapy. The cDNA is used for recombinant expression of iENTP and as a source of oligonucleotides (diagnostic primers and probes, ribozymes and antisense sequences). Antibodies raised against iENTP are used for detection of the protein by usual immunoassays and as (ant)agonists of iENTP activity. This sequence represents the human ENT1 protein and is used for comparison with the human NBMPR-iENTP protein (AAB15517).

Sequence 456 AA;

Query Match 15.0%; Score 343.5; DB 21; Length 456;
 Best Local Similarity 24.3%; Pred. No. 7e-27;
 Matches 112; Conservative 77; Mismatches 212; Indels 59; Gaps 13;

Qy 18 PRDXYNYVWLVGVGVLPLWNNFTTAPYVYVWFKPDGVE---TWYSKFMGSLT 74
 Db 7 pddrykavllfmgllgllpwnffmt-atqyftndmsqnsvltaelskdaqasaa 65

Qy 75 IGSOLP--NASINVFNLIIAGLLIYVFEA-----PVCENIVN----LTITLT 117
 Db 66 paalpernsisaifnnvmticamlplllfytlnsfhlqrlpdsrvllgsvalllvfl 125

Qy 118 LVIVLEPTDSMSFFFWWTLGMATSNFNSGLYNSVYGVGDPFHTYIGALLIGNTCG 177
 Db 126 tailkvqldalp-ffvitmkivlnsfagilqsglsfagllpasytapimgsgqglag 184

Qy 178 LLITVWIGVTFYFNDEPKLVAIVFGISVLVLCALFFYKQDPYHYHQ----- 231
 Db 185 ffasvami-calasgselsesafyfitacaviiltiicylgplrlefyryyqqlieg 243

Qy 232 -----KGMEIREKAETDRSPSILWTT-----FTNCYQLFNWVFCFAVT 271
 Db 244 geqetkidlsgkeepragkeesgvsvnsqptneshsikaiklnisvafsvcfifit 303

Qy 272 LTIPVVMVTTRGDSGLFNKIMSENDEIYTLTSLVFNLFNFAAIGSIVASKIHWP--TP 329
 Db 304 igmfpaevtvekvsslag-----sstweryfipvscfltnifdwgrsltavfmpgkds 358

Qy 330 RYLFAILRALFIPFFFCNRYVOTRAY-PVFESTDIFVIGGIAMSFSGHYSALAMG 388
 Db 359 rwlpslvarlrvfpilllcn--ikpryrltvfhdawfiffmaafasngylaslcmc 416

Qy 389 YTPNVPFVSHYSRAAQLSVCTLMWGLLTGGLPVPIEHFV 428
 Db 417 fgpkkvpaetaetagaiafflciglalgvafslfraiv 456

RESULT 5

AY82285
 ID AY82285 standard; Protein; 475 AA.
 AC AY82285;

DT 19-JUN-2000 (first entry)

DE Human ENT1 receptor SEQ ID NO:1.

Human; ENT1; ENTR1; immunosuppressant; vasotropic; thrombolytic;

KW cytostatic; hypotensive; antiinflammatory; analgesic; anticoagulant;
 KW nucleoside transporter; ischaemia; cerebral embolism; malignant tumour;
 KW organ transplant rejection; renitis; pancreatitis; hypertension;
 KW analgesic; blood platelet coagulation inhibitor.

OS Homo sapiens.

PN WO200012350-A1.

XX 09-MAR-2000.

XX 26-AUG-1999; 99WO-JP04602.

XX 27-AUG-1998; 98JP-0241248.

PA (KYOW) KYOWA HAKKO KOGYO KK.

PI Miyaaji H, Mimura H, Kambe M, Nakagawa S;

XX WPI; 2000-256588/22.

DR N-PSDB; AAA08027.

XX Novel nucleoside transporter polypeptide, useful in drug development
 PT for agents in treating e.g. ischaemia, cerebral embolism, rejection in
 PT organ transplant, malignant tumors, renitis, pancreatitis and
 PT hypertension

PS Claim 1; Page 70-74; 94pp; Japanese.

XX The present invention describes human and rat ENT1 receptors (ENTR1),
 CC with some amino-acids deleted, substituted or added, but which retain
 CC nucleoside transport activity, and can be used in drug development for
 CC agents in treating e.g. ischaemia, cerebral embolism, rejection in organ
 CC transplant, malignant tumors, renitis, pancreatitis and hypertension.
 CC The ENTR1 proteins are nucleoside transporters which have vasotropic,
 CC immunosuppressant, thrombolytic, cytostatic, hypotensive, analgesic,
 CC antiinflammatory and anticoagulant activities. The transporter proteins,
 CC their DNA, oligonucleotides, transformants and antibodies can be used
 CC in drug development for agents in treating ischaemia, cerebral embolism,
 CC rejection in organ transplant, malignant tumors, renitis, pancreatitis
 CC and hypertension, and as analgesics, or blood platelet coagulation
 CC inhibitors, as well as agents to reduce side-effects during
 CC chemotherapy. The present sequence represents the human ENTR1 protein.

XX Sequence 475 AA;

Query Match 14.3%; Score 327.5; DB 21; Length 475;
 Best Local Similarity 24.4%; Pred. No. 3.3e-25;
 Matches 116; Conservative 83; Mismatches 202; Indels 75; Gaps 15;

Qy 7 SNTYAVEQEA-----PRDKYNYVWLVGVGVLPLWNNFTTAPYV 52
 Db 21 ssslraddalleklldrrppglrpedrfcgtlyiffslgslpwnffit-akeywm 79

Qy 53 NYWFK-----PDGVETWYSKFMGSLTIGSOLPNASINVFNLIIAGLLIYR 100
 Db 80 ---fklrnspsatgedpsd--ilnyfesylavastvpsmclvanflvnrvavhir 134

Qy 101 VFAPVCNIVNLTILILVIVLEP--TEDSMW---PFWTLGMATSNFNSGLYNSVY 155
 Db 135 vla-----sltvilaifmvtalvkvtftswtrgffavtvcmvilsgastvfsssi 187

Qy 156 GVGSGPHTYIGALLIGNNICGLLTIVVKGIGVTFYNDEPKLVAIVFGISVLVLCAL 215
 Db 188 gmtsgfpmrnsqalissgmggtvsavasi-vdlaasdsrnsalafitafivlcmg 246

Qy 216 ALFFITRKQDFYHY-----HHQKMEIREKAETDRSPSI-----LWTFN 256
 Db 247 lyllslrlyaryymrpvlaahvfsgge--elpqdsapsvasrfdshpplrk 304

Qy 257 CYGOLFNVWFCFATVLTIFPVMMVTYTRGDSGLFNKIMSE--NDEIYTLTSLVFNLF 314

Db 305 taslgfctvtyvflltsliypavct-----nieslnksgslwttkfflptlttlllynfad 359

QY 315 ATGSTVASKIHWTP--RYLKEFAIILRALFIPFEFFFCNVRVQTRAYPVFFESTDIFVIGG 372

Db 360 lcgrgtawiqvpgnpskalpofvrlrtclplflvclnqprvhlktvfgsdvypalls 419

QY 373 IAMSFSHGYSALAMGYTPNVVPSHYSRFAAOLSVCTLMVGLLTGLWLPVVIHEFV 428

Db 420 silglsgylstlallypkivprelaeatgvmsfyvclgltsacstllivhii 475

RESULT 6

AAW83965

ID AAW83965 standard; Protein: 456 AA.

XX AC AAW83965;

XX 08-FEB-1999 (first entry)

XX Equilibrative nucleoside transport protein (iENTP).

XX Equilibrative nucleoside transport protein; iENTP: NBMPR: transport;

KW nitrobenzylmercaptapurine riboside; antiviral; antitumour; screening;

KW inhibitor; drug; knockout mice; cancer; chemotherapy; gene therapy;

KW adenosine deaminase; factor VIII.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Domain 13..27

FT /note= "transmembrane domain TM1"

FT Domain 71..93

FT /note= "transmembrane domain TM2"

FT Domain 100..117

FT /note= "transmembrane domain TM3"

FT Domain 125..144

FT /note= "transmembrane domain TM4"

FT Domain 162..185

FT /note= "transmembrane domain TM5"

FT Domain 194..214

FT /note= "transmembrane domain TM6"

FT Domain 293..312

FT /note= "transmembrane domain TM7"

FT Domain 326..341

FT /note= "transmembrane domain TM8"

FT Domain 361..379

FT /note= "transmembrane domain TM9"

FT Domain 393..413

FT /note= "transmembrane domain TM10"

FT Domain 432..452

FT /note= "transmembrane domain TM11"

XX WO9846749-A1.

XX 22-OCT-1998.

XX 10-APR-1998; 98WO-US07283.

XX 09-APR-1998; 98US-0058389.

PR 11-APR-1997; 97US-0838845.

XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

XX PA Belt JA, Crawford CR, Patel DH;

XX PI WPI; 1998-594576/50.

DR N-PSDB; AAW69698.

XX New equilibrative nucleoside transport protein insensitive to

PT nitrobenzylthio-purine riboside - useful for, e.g. identifying

PT specific therapeutic nucleoside analogues and in gene therapy to

XX protect transduced cells against ablative chemotherapy

PS Claim 13; Pages 71-72; 114pp; English.

XX This represents an equilibrative nucleoside transport protein (iENTP),

CC which is insensitive to nitrobenzylmercaptapurine riboside (NBMPR). Cells

CC transformed with a construct containing the iENTP nucleic acid can be

CC used to produce the protein recombinantly. iENTP is used to identify

CC specific ligands (particularly antiviral and antitumour nucleoside

CC analogues that are preferentially transported into cells) and to raise

CC antibodies. Cells in which iENTP provides all available transport

CC activity are used: (a) to identify permeants of iENTP and (b) to screen

CC specific inhibitors of iENTP (potential drugs). Fragments of the iENTP

CC nucleic acid are used, as probes, primers, antisense molecules, and

CC ribozymes for therapy or diagnosis, and knockout mice in which both

CC alleles encoding iENTP contain an inactivating defect are also useful for

CC drug screening. Cells that have been transduced with iENTP nucleic acid

CC ex vivo are used particularly for cancer chemotherapy. Vectors in which

CC the iENTP gene is linked to a heterologous gene (e.g. encoding adenosine

CC deaminase and factor VIII) can be used for gene therapy.

XX Sequence 456 AA:

Query Match 14.2%; Score 326.5; DB 19; Length 456;

Best Local Similarity 22.9%; Pred. No. 4e-25;

Matches 106; Conservative 71; Mismatches 200; Indels 85; Gaps 11;

QY 18 PRDKYNIYVYLVILVCGVLLPNMFIITAPEY-----VNYWFKPDGVE 62

Db 7 prdsyhlvgisffllgltclpwnffitaipfqlarlagagnstariltnhgpedaf- 65

QY 63 TWYSKEFGSLTIGSOLPNASINVFNLFIAGIYRVFAPVCFNIVNLTIILIVIVL 122

Db 66 -----nfnmwvllsqpllllftllnsflyqcvpetvrlg---slalillfaltaal 116

QY 123 EPTEDSMWFFWWTGLMATSFNSNGLYENSVYGVGGDPHYIYIGALLIGNNICGLLTV 182

Db 117 vkvdmspgpffsitmasvcfinsfavlqgsifglgtmpstystiflsgqglagifaa 176

QY 183 VKI-----GVTYFLNDEPKLVAIVYF---GISLVILLVCAIALFFITKODFY---HYHHQ 231

Db 177 amllsmasgv-----daetsalgyfitpcvgilmsivcylslphikfarylankssga 230

QY 232 KGMEIREKA-----ETDRPSPSILMTTFTNC 257

Db 231 qaqeletkaellqsdengipspqkvalltdldekepesepepdkpqpksvftvfkil 290

QY 258 YGQLFNWVWFCFAVTLTIFPVMMVTVTTRGDSGFLNKIMSENDELYTLTSLFVFNLEAAIG 317

Db 291 witaiclvlvftvtlsvpaitamvtssts-----pgkwsqffnfpiccllfnmdwig 344

QY 318 SIVASKIHWTP--TPRYLKEFAIILRALFIPFEFFFCNVRVQTRAYPVFFESTDIFVIGGIAM 375

Db 345 rsitsyflwpdedsrllpllvclrlfvpplmclhvpqrsr-lpifpdaqdayfitmlif 403

QY 376 SFSHGYSALAMGYTPNVVPSHYSRFAAOLSVCTLMVGLLTG 417

Db 404 avsnqylvsitmcclapqrqlpherevagalmftfialgiscg 445

RESULT 7

AAW69558

ID AAW69558 standard; Protein: 456 AA.

XX AC AAW69558;

XX 13-OCT-1998 (first entry)

XX Human equilibrative nucleoside transporter 2.

XX Human; equilibrative nucleoside transporter; hENT1; hENT2; rENT1;

KW rENT2; coronary; cerebrovascular anoxia; viral infection; cancer.

XX OS Homo sapiens.

XX WO9829437-A2.
 XX 09-JUL-1998.
 XX 30-DEC-1997; 97WO-IB01657.
 XX 03-NOV-1997; 97US-0064004.
 XX 30-DEC-1996; 96US-0034083.
 XX (UYAL-) UNIV ALBERTA.
 XX (UYLE-) UNIV LEEDS.
 XX Baldwin SA, Cass CE, Young JD;
 XX WPI: 1998-388035/33.
 XX N-PSDB: AAV40279.
 XX Newly isolated equilibrative nucleoside transporter protein(s) and
 PT gene(s) - used to develop products for treating disorder(s)
 PT associated with the transporter(s) and for use with nucleoside
 PT drug(s)
 XX
 PS Claim 5; Fig 20; 97pp; English.
 XX
 CC The present sequence represents a substantially purified equilibrative
 CC nucleoside transporter (ENT), human ENT2 (hENT2). ENTs can transport a
 CC variety of purines and pyrimidines, including adenosine, uridine,
 CC guanosine, inosine, formycin B, tubercidin, and thymidine. ENTs are
 CC bidirectional, they transport a suitable permeant both into and out of
 CC cells. ENTs can be used as a tool for the development of new nucleoside
 CC drugs. Products from the present invention can be used for treating a
 CC subject having a disorder associated with an ENT. They can also be used
 CC with nucleoside drugs in the treatment of e.g. coronary or
 CC cerebrovascular anoxia, viral infection or cancer. The products (e.g.
 CC antibodies and oligonucleotides) hybridising to nucleic acid sequences
 CC encoding ENTs) can also be used for detection and drug screening.
 XX
 SQ Sequence 456 AA;

Query Match 14.28; Score 326.5; DB 19; Length 456;
 Best Local Similarity 22.9%; Pred. No. 4e-25;
 Matches 106; Conservative 71; Mismatches 200; Indels 85; Gaps 11;

QY 18 PRDKNYIVMLVILVGVLLPWNMTTIAPEYY-----VNWFKPDGVE 62
 DB 7 prdshyhgvisffilgtltpwnffitaipyfqlarlagagnstarlsthtgpedaf- 65
 QY 63 TWYSKEFMGSLTIGSQLPNASINVENFLFIITAGPLIYRVFAPVCFNIVNLTIILVIVL 122
 DB 66 -----nfnmwvllsqplllftllnsflyqcvpetvrlg----sllaillfaltaal 116
 QY 123 EPTEDSMWFFWTLGMATSNFNSGLYENSIVGVGDFPHYTYGALLIGNNICGLITV 182
 DB 117 vkvdmspgpffsiltmasvcfinsavlgqslfglgtmpststlflsggglagifaaal 176
 QY 183 VKI-----GVTVFLNDPKLVAIVVF---GISVLVLVCAIALFFITKQDPY---HVHHQ 231
 DB 177 amllismasgv-----daetsalgyfitpyvgilmsivcylsiphlkfarylankssqa 230
 QY 232 KGMIREKA-----ETDRSPSILMTTPTNC 257
 DB 231 qaqeletkaellsgdengipsspqkvaltdldlekepeseppdepqpgkpsvtfvqki 290
 QY 258 YGQLFNWFCFAVTLTFPVNMTVTRGDSGLFNKINSENDEIYTLTSLVFNLFRAIG 317
 DB 291 wltalcivlvfvcisvipaltamvts-----Pgwksqffnplccflfnimdwlg 344
 QY 318 SIVASKIHP--TPRYLKFAILLRALRPIPFPPFCNVRQVTRAYPVFFESTDIFVIGGIAM 375
 DB 345 rslcysflwpderrllpllvclrfvflvpfmlchvqrsr-lpillfqdayfitfmlif 403

QY 376 SFSHCYLSALAMGYTPNVVPSHYSRFAAQLSVCTLMVGLLTG 417
 DB 404 avsgylvsltmclaprvlpherevagalmntfflalglscg 445

RESULT 8
 AAB15517
 ID AAB15517 standard; Protein: 456 AA.
 XX
 AC AAB15517;
 XX
 DT 14-FEB-2001 (first entry)
 XX
 DE Human NBMPR-iENTP protein.
 XX
 KW Antiviral; antitumour; NBMPR-iENTP; nitrobenzylmercaptopyrineriboside;
 KW insensitive, equilibrative nucleoside transporter protein; cancer;
 KW plasma membrane; facilitated diffusion; gene therapy; primer; probe.
 XX
 OS Homo sapiens.
 XX
 PN US6130065-A.
 XX
 PD 10-OCT-2000.
 XX
 XX 09-APR-1998; 98US-0058389.
 PF
 XX 11-APR-1997; 97US-0043659.
 PR
 XX (SUUD-) ST JUDE CHILDREN'S RES HOSPITAL.
 PA
 XX Crawford CR, Belt JA;
 PI
 XX WPI: 2000-637839/61.
 XX
 DR N-PSDB: AAA95722.
 DR
 XX New nucleic acid encoding nucleoside transporter, useful for drug
 PT screening to identify antitumour and antiviral agents and for gene
 PT therapy
 PT
 XX
 PS Claim 1; Fig 2A-B; 55pp; English.
 XX
 CC This sequence represents nitrobenzylmercaptopyrineriboside-insensitive,
 CC equilibrative nucleoside transporter protein (NBMPR-iENTP) isolated from
 CC Hela cells. The protein transports nucleosides across the plasma membrane
 CC by a facilitated diffusion process and is involved in the proliferative
 CC response. The transporter is useful in screening to identify natural
 CC nucleoside permeants and/or their inhibitors or analogues, potential
 CC therapeutic agents, also for studies on protein structure and mechanism.
 CC Cells that express the iENTP as the only transporter protein are used
 CC for drug screening (especially to identify antitumour and antiviral
 CC nucleoside analogues), in chemotherapy of cancer and for selective
 CC expression of heterologous genes for gene therapy. The cDNA is used
 CC for recombinant expression of iENTP and as a source of oligonucleotides
 CC (diagnostic primers and probes, ribozymes and antisense sequences).
 CC Antibodies raised against iENTP are used for detection of the protein
 CC by usual immunoassays and as (ant)agonists of iENTP activity.
 XX
 SQ Sequence 456 AA;

Query Match 14.28; Score 326.5; DB 21; Length 456;
 Best Local Similarity 22.9%; Pred. No. 4e-25;
 Matches 106; Conservative 71; Mismatches 200; Indels 85; Gaps 11;

QY 18 PRDKNYIVMLVILVGVLLPWNMTTIAPEYY-----VNWFKPDGVE 62
 DB 7 prdshyhgvisffilgtltpwnffitaipyfqlarlagagnstarlsthtgpedaf- 65
 QY 63 TWYSKEFMGSLTIGSQLPNASINVENFLFIITAGPLIYRVFAPVCFNIVNLTIILVIVL 122
 DB 66 -----nfnmwvllsqplllftllnsflyqcvpetvrlg----sllaillfaltaal 116

PR	17-SEP-1998;	98US-0100688;
PR	17-SEP-1998;	98US-0100684;
PR	17-SEP-1998;	98US-0100710;
PR	17-SEP-1998;	98US-0100711;
PR	17-SEP-1998;	98US-0100919;
PR	17-SEP-1998;	98US-0100930;
PR	18-SEP-1998;	98US-0100848;
PR	18-SEP-1998;	98US-0100849;
PR	18-SEP-1998;	98US-0101014;
PR	18-SEP-1998;	98US-0101068;
PR	18-SEP-1998;	98US-0101071;
PR	22-SEP-1998;	98US-0101279;
PR	23-SEP-1998;	98US-0101471;
PR	23-SEP-1998;	98US-0103472;
PR	23-SEP-1998;	98US-0103473;
PR	24-SEP-1998;	98US-0101741;
PR	24-SEP-1998;	98US-0101743;
PR	23-SEP-1998;	98US-0101475;
PR	23-SEP-1998;	98US-0101476;
PR	23-SEP-1998;	98US-0101477;
PR	23-SEP-1998;	98US-0101478;
PR	23-SEP-1998;	98US-0101730;
PR	23-SEP-1998;	98US-0103240;
PR	29-SEP-1998;	98US-0102307;
PR	29-SEP-1998;	98US-0102307;
PR	29-SEP-1998;	98US-0102331;
PR	29-SEP-1998;	98US-0102331;
PR	30-SEP-1998;	98US-0102484;
PR	30-SEP-1998;	98US-0102487;
PR	30-SEP-1998;	98US-0102570;
PR	30-SEP-1998;	98US-0102571;
PR	01-OCT-1998;	98US-0102684;
PR	01-OCT-1998;	98US-0102687;
PR	02-OCT-1998;	98US-0102965;
PR	06-OCT-1998;	98US-0103258;
PR	06-OCT-1998;	98US-0103449;
PR	07-OCT-1998;	98US-0103314;
PR	07-OCT-1998;	98US-0103315;
PR	07-OCT-1998;	98US-0103328;
PR	07-OCT-1998;	98US-0103395;
PR	07-OCT-1998;	98US-0103401;
PR	08-OCT-1998;	98US-0103633;
PR	08-OCT-1998;	98US-0103678;
PR	08-OCT-1998;	98US-0103679;
PR	14-OCT-1998;	98US-0103711;
PR	22-OCT-1998;	98US-0105104;
PR	22-OCT-1998;	98US-0105266;
PR	26-OCT-1998;	98US-0105693;
PR	20-OCT-1998;	98US-0105000;
PR	20-OCT-1998;	98US-0105002;
PR	21-OCT-1998;	98US-0105104;
PR	22-OCT-1998;	98US-0105109;
PR	22-OCT-1998;	98US-0105266;
PR	26-OCT-1998;	98US-0105693;
PR	26-OCT-1998;	98US-0105694;
PR	27-OCT-1998;	98US-0105807;
PR	27-OCT-1998;	98US-0105881;
PR	27-OCT-1998;	98US-0105882;
PR	27-OCT-1998;	98US-0106062;
PR	28-OCT-1998;	98US-0106023;
PR	28-OCT-1998;	98US-0106029;
PR	28-OCT-1998;	98US-0106030;
PR	28-OCT-1998;	98US-0106032;
PR	28-OCT-1998;	98US-0106033;
PR	28-OCT-1998;	98US-0106033;
PR	28-OCT-1998;	98US-0106178;
PR	29-OCT-1998;	98US-0106248;
PR	29-OCT-1998;	98US-0106384;
PR	29-OCT-1998;	98US-0106384;
PR	30-OCT-1998;	98US-0108500;
PR	30-OCT-1998;	98US-0108464;
PR	03-NOV-1998;	98US-0106856;
PR	03-NOV-1998;	98US-0106902;

QY 7 SNTYAVEQEAR-----PRDKYNIYVWLVLVGVGLLPWNMFITIAPEYV 52
 Db 21 sssiradqealleklldrpppgldrpedrfgctyilffslgslgllpwnffit-akeywm 79
 QY 53 NYWFK-----PDGVETWYSKEFGMSLIGSOLPNASINVFNLFIITAGPLIYR 100
 Db 80 ---fklrnssspatgedpegsd--ilnyfesylavastvpsmlclvanflilnrvavhir 134
 QY 101 VFAPVCFNIVNLITILIVILEP--TEDSMW--FEWVTLGMATSNFNSGLYNSVY 155
 Db 135 vla-----siltvllaifmvtalvkdtsrgrffavtvcvmlisgastvissly 187
 QY 156 GVGDFPHYICALLIGNNICLLITVVKIGVTVFLNDEPKLVAIVYFGISLVILLVCAI 215
 Db 188 gmtsgfmrnsqalisgagmtgtsavasl-vdlaassdvnsalafiltatfivlcmg 246
 QY 216 ALFITTKDFVHY-----HHQKGEIREKAETDRPSI-----LWTFTN 256
 Db 247 lylilslrleyarymrvpvaahvfsgge--elpqdsisapsvasrfdidshptpplrk 304
 QY 257 CYGOLFNNWCFATLRTFPVMTVTRGDSGLNKIMSE--NDEIYTLTSLFLVNLFA 314
 Db 305 taslgfcvtyvffitsllypavct-----nieslnksgslwtckfplltflllynfad 359
 QY 315 AIGSIVASKIHWTP--RYLKFAILLRALFIPFFFFCNRYVOTRAYPVFFESTDIFVIGG 372
 Db 360 lcgqltawigvppnskalpfgvllrtcltclpflvclnyqrvhlktvfgsdvypalls 419
 QY 373 IAMSFSHGYSALAMGYTPNVVPSHYSRFAAQLSVCTLMVGLLTGGLWPPVVIHFV 428
 Db 420 slilgslngyistlalygpkivprelaeatgvmsfyvclgltlgsacstllvhli 475

RESULT 11

AAW69557
 ID AAW69557 standard; Protein: 456 AA.

XX AC AAW69557;

XX DT 13-OCT-1998 (first entry)

XX DE Rat equilibrative nucleoside transporter 2.

XX KW Rat; equilibrative nucleoside transporter; hENT1; hENT2; rENT1;
 XX KW rENT2; coronary; cerebrovascular anoxia; viral infection; cancer.
 XX OS Rattus sp.

XX PN W09829437-A2.

XX PD 09-JUL-1998.

XX PF 30-DEC-1997; 97WO-IB01657.

XX PR 03-NOV-1997; 97US-0064004.

XX PR 30-DEC-1996; 96US-0034083.

XX PA (UYAL-) UNIV ALBERTA.

XX PA (UYLE-) UNIV LEEDS.

XX PI Baldwin SA, Cass CE, Young JD;

XX DR WPI: 1998-388035/33.

XX DR N-PSDB: AAV40278.

XX PT Newly isolated equilibrative nucleoside transporter protein(s) and
 PT gene(s) - used to develop products for treating disorder(s)
 PT associated with the transporter(s) and for use with nucleoside
 PT drug(s)

XX PS Claim 5; Fig 14; 97pp; English.

XX XX

CC The present sequence represents a substantially purified equilibrative
 CC nucleoside transporter (ENT), rat ENT2 (rENT2). ENTs can transport a
 CC variety of purines and pyrimidines, including adenosine, uridine,
 CC guanosine, inosine, formycin B, tubercidin, and thymidine. ENTs are
 CC bidirectional, they transport a suitable permeant both into and out of
 CC cells. ENTs can be used as a tool for the development of new nucleoside
 CC drugs. Products from the present invention can be used for treating a
 CC subject having a disorder associated with an ENT. They can also be used
 CC with nucleoside drugs in the treatment of e.g. coronary or
 CC cerebrovascular anoxia, viral infection or cancer. The products (e.g.
 CC antibodies and oligonucleotides hybridising to nucleic acid sequences
 CC encoding ENTs) can also be used for detection and drug screening.
 XX
 SQ Sequence 456 AA;

Query Match 14.0%; Score 320.5; DB 19; Length 456;
 Best Local Similarity 23.5%; Pred. No. 1.7e-24;
 Matches 107; Conservative 76; Mismatches 200; Indels 73; Gaps 13;

QY 18 PRDKYNIYVWLVLVGVGLLPWNMFITIAPEYVYVWFKPDGVETWYSK-----EF 69
 Db 7 prdshlvigisffligtllpwnffitaiipfgrlagtussaeptshntspdtfnf 66
 QY 70 MGSUTIGSQLPNASINVFNLFIITAGPLIYRVFAPVCFNIVNLFIITILIVILEPTEDSM 129
 Db 67 nnwvllsqlllftllnsflyqcipesvrlg----sllaillfaltaalvkvdls 122
 QY 130 SWFVWTLGMATSNFNSGLYNSVYGVGGDPHYTIGALLIGNNICGL-----LITVVK 184
 Db 123 glffsitmasvfwinsfcavlgqslfqlgmpstystlflsggqlagifaalamltsla 182
 QY 185 IGVYFYLNDEPKLVAIVYF--GISLVILLVCAIAL-----PFIKQ----- 223
 Db 183 sgv-----qbqtsalgyfitpcvgillisiicyslphlkfaryytkkqapvqeletk 236
 QY 224 -DFYHYHHQKMEIR-----EK-----AETDRPSPSILMTFTTCYQGLFNV 264
 Db 237 aellgadekngipvpqagptldldpekeleleleqkpgkpsvfvfrkiwitalci 296
 QY 265 WFCFAVLTITFP-VMMVTVTRGDSGLFNKIMSENDEIYTLTSLFLVNLFAIGSIVASK 323
 Db 297 vlvftvtlsvfpaitamvttssns-----pgkwsqffnpiiccflfnvmdwlgrrls 350
 QY 324 IHWP--TPRYLKFAILLRALFIPFFFFCNRYVOTRAYPVFFESTDIFVIGGIAWSFSHG 381
 Db 351 flwpdedsqllpllvclrfllvplfmlchvppqr-lpifwqdayfitfmlfaisngy 409
 QY 382 LSALAMGYTPNVVPSHYSRFAAQLSVCTLMVGLLTG 417
 Db 410 fvsltmclaprqviphervagalmfllalglscg 445

RESULT 12

AAW69556

ID AAW69556 standard; Protein: 457 AA.

XX AC AAW69556;

XX DT 13-OCT-1998 (first entry)

XX DE Rat equilibrative nucleoside transporter 1.

XX KW Rat; equilibrative nucleoside transporter; hENT1; hENT2; rENT1;
 XX KW rENT2; coronary; cerebrovascular anoxia; viral infection; cancer.

XX OS Rattus sp.

XX PN W09829437-A2.

XX PD 09-JUL-1998.

XX PF 30-DEC-1997; 97WO-IB01657.

```

XX 03-NOV-1997; 97US-0064004.
PR 30-DEC-1996; 96US-0034083.
XX
XX (UYAL-) UNIV ALBERTA.
PA (UYLE-) UNIV LEEDS.
XX
XX Baldwin SA, Cass CE, Young JD;
PI
XX
XX WPI: 1998-388035/33.
DR N-PSDB; AAV40277.
XX
XX Newly isolated equilibrative nucleoside transporter protein(s) and
PT gene(s) - used to develop products for treating disorder(s)
PT associated with the transporter(s) and for use with nucleoside
PT drug(s)
XX
XX Claim 5; Fig 9; 97pp; English.
XX
XX The present sequence represents a substantially purified equilibrative
CC nucleoside transporter (ENT), rat ENT1 (rENT1). ENTs can transport a
CC variety of purines and pyrimidines, including adenosine, uridine,
CC guanosine, inosine, formycin B, tubercidin, and thymidine. ENTs are
CC bidirectional, they transport a suitable permeant both into and out of
CC cells. ENTs can be used as a tool for the development of new nucleoside
CC drugs. Products from the present invention can be used for treating a
CC subject having a disorder associated with an ENT. They can also be used
CC with nucleoside drugs in the treatment of e.g. coronary or
CC cerebrovascular anoxia, viral infection or cancer. The products (e.g.
CC antibodies and oligonucleotides) hybridising to nucleic acid sequences
CC encoding ENTs) can also be used for detection and drug screening.
XX
XX Sequence 457 AA;
SQ
Query Match 13.1%; Score 301; DB 19; Length 457;
Best Local Similarity 22.3%; Pred. No. 1.7e-22;
Matches 107; Conservative 84; Mismatches 191; Indels 98; Gaps 17;
QY 18 PROKYNIVWLTVLGVLLPNWMTTIAPEYVYVWFKPDGV-----ETWYSKFFMGS 72
Db 7 pqrqkxwllfvlgtltpwnffit-atqytsrlntsqnivtngscstealad 65
QY 73 LTGSGOLPNASINNVNLFLLIAGPLIYRVFAPVCFN-----IVNLTII 115
Db 66 psvslparssalsafnvnmtlcampllif--tclnsflhqkvsgslrlgslalllvf 123
QY 116 LILVIVLEPTEDSMWFFWTLGCMATSYNSGLYNSVYGVGDPPHTYVIGALLIGNNI 175
Db 124 lvtativkvqmdals-fftiitmikivlinsfgallqasifglagvipanytapimgsgql 182
QY 176 CGLLITVVKI-----GVTFLNDEPKLVAIVFGISLVILLVCAIALFFFTKQ 223
Db 183 agftsvamicavsgskisesafyfit---acavv-----ilaiclyalpwm--- 229
QY 224 DFYHHQ-----KGMETR-EKAETDRSPSIL----- 250
Db 230 efyrhylvlqnlagpaegtkldiisegeprgreesvgpgpnsllpanrnqskailksl 289
QY 251 WTTFTNCGOLFNVWFCFVTLTIFPVMVTVTRGDSGLFKIMSENDEITYTLTSLVLF 310
Db 290 wvl-----alsvcfiftvtlglfp---avtaevessiaagtspwknc-yflpvactlnf 338
QY 311 NLEFAAGISIVASKIHP--TPRYLKFAILRALFIFPFFFCNRYVOTRAYPVVFESTDIF 368
Db 339 nvdwlgrrsltalcmppgqdsrwlplvacrvvfiplllmclcnvk-qhhyllpslfdkdwf 397
QY 369 VIGGIAMSFSGHLSALAMGYTPNVVPSHSYRFAALSVCTLMVGLLTGGLWPVVIEHFV 428
Db 398 itfmaafatsngylaslcmcfqgkpkvpaetaetagnimsfflclglalgavslfllralv 457

```

RESULT 13

```

AAW64550
ID AAW64550 standard; Protein; 373 AA.
XX
AC AAW64550;
XX
XX 21-OCT-1998 (first entry)
XX
XX Human fibrosarcoma cell line HT-1080 clone HP10235 protein.
XX
XX Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW differentiation; immune system; stimulator; suppressor; regulator;
KW haematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour.
XX
OS Homo sapiens.
XX
XX WO9821328-A2.
XX
XX 22-MAY-1998.
XX
XX 07-NOV-1997; 97WO-JP04056.
XX
XX 13-NOV-1996; 96JP-0301429.
XX
XX (PROT-) PROTEGENE INC.
XX (SAGA ) SAGAMI CHEM RES CENTRE.
XX
XX Kato S, Kobayashi M, Sekine S, Yamaguchi T;
XX WPI: 1998-297932/26.
XX DR N-PSDB; AAV49582, AAV49583.
XX
XX Human protein having transmembrane domain - useful for, e.g.
XX research and nutrition
XX
XX Claim 1; Page 108-109; 205pp; English.
XX
XX AAW64534-W64558 represent human proteins containing a transmembrane
CC domain. These proteins can be used for, e.g. research and nutrition, and
CC may have cytokine and cell proliferation/differentiation, immune
CC stimulating/suppressing, haematopoiesis regulating, tissue growth,
CC activin/inhibin, chemotactic/chemokinetic, haemostatic and thrombolytic,
CC receptor/ligand, anti-inflammatory or tumour inhibition activity.
XX
XX Sequence 373 AA;
SQ
Query Match 11.9%; Score 272.5; DB 19; Length 373;
Best Local Similarity 23.6%; Pred. No. 1.1e-19;
Matches 91; Conservative 66; Mismatches 187; Indels 41; Gaps 9;
QY 73 LTGSGOLPNASINNVNLFLLIAGPLIYRVFAPVCFNIVNLTIIILVIVLEPTEDSMWF 132
Db 1 mticamlplllfynsflhqripqsvrlgsl---vallvflitalivkvqldalp-f 56
QY 133 FWTGLMATSYNSGLYNSVYGVGDPPHTYVIGALLIGNNICGLLITVVKIVGYFLN 192
Db 57 fvitmikivlinsfgallqasifglaglipasytapimgsggagffasvami-caiasg 115
QY 193 DEPKLVAIVFGISLVILLVCAIALFFITKQDEYHHQ-----KGME 235
Db 116 selseafgyfitaacavviltiiclyglprletfryyqqllegpgeqetkldliskge 175
QY 236 IREKAETDRSPSILWTT-----FTNCGOLFNVWFCFVTLTIFPVMVTVTRGD 286
Db 176 pragkeesgvsvnsopntshesikaillnislavsfvcfiftitigmfpavtvekssi 235
QY 287 SGLFNKIMSENDEITYTLTSLVFLNFAAGISIVASKIHP--TPRYLKFAILRALFIP 344
Db 236 ag-----sstweryfipvcfltnifdlgrsltavfmwpgkdsrwlpslvarlvfp 290
QY 345 FFFFCNRYVOTRAY-PVVFESTDIFVIGGIAMSFSGHLSALAMGYTPNVVPSHSYRFAA 403

```


CC involved in the proliferative response. The transporter is useful in
 CC screening to identify natural nucleoside permeants and/or their
 CC inhibitors or analogues, potential therapeutic agents, also for studies
 CC on protein structure and mechanism. Cells that express the iENTP as the
 CC only transporter protein are used for drug screening (especially to
 CC identify antitumor and antiviral nucleoside analogues), in chemotherapy
 CC of cancer and for selective expression of heterologous genes for gene
 CC therapy. The cDNA is used for recombinant expression of iENTP and as
 CC a source of oligonucleotides (diagnostic primers and probes, ribozymes
 CC and antisense sequences). Antibodies raised against iENTP are used for
 CC detection of the protein by usual immunoassays and as (antagonists of
 CC iENTP activity. This sequence represents the human HNP36 protein and
 CC is used for comparison with the human NBMPR-iENTP protein (AAB15517).

xx
 SQ Sequence 326 AA;

Query Match 9.3%; Score 214; DB 21; Length 326;
 Best Local Similarity 22.0%; Pred. No. 1e-13;
 Matches 71; Conservative 52; Mismatches 139; Indels 60; Gaps 8;

QY 143 INFSGLYENSVGVGDFPHYIGALLIGNNIGCLLITVVKI-----GVITYFLNDEPKL 197
 Db 7 insfsavlgqsfqglqmpstystlflsgqglagifaafamllsmasgv-----daet 60
 QY 198 VAIYVF---GISLVILLVCAIALEFFITKQDFY---HYHHOKGMEIREKA----- 240
 Db 61 saigyfitpyvgilmsivcylslphlkfarylankssqqaqaqelektkaellqsdengip 120
 QY 241 -----ETDRSPSILMTTNTCYGQLFNWVFCFAVTLTTFPV 277
 Db 121 sspqkvaltidldlekesepdepqpgkpsvftvfqkiwltalcivlvftvltlsvfpa 180
 QY 278 MMTVTTRGDSGLNKINSENDEIYTLTSLFVNLFRAIGSIVASKIHP--TPRYLKFA 335
 Db 181 itamvtsts-----pgkwsqffnppiccfllfnmdwlgrrsltsyflwpdedsrllp 234
 QY 336 IILRALFPFFFCNRYVQTRAVPVFFESTDIFVIGGIAMSFSGHLSALAMGYTPNVVP 395
 Db 235 vclrfllvpfmlchvprsr-lpilfpqdayfitmllfavsngyivslmtclaprql 293
 QY 396 SHYSRFAAQLSVCTLMVGLLTG 417
 Db 294 pherevagalmftflalglscg 315

Search completed: February 27, 2002, 16:59:32
 Job time: 199 sec